

# TET proteins and DNA modifications in epigenetic regulation

Dissertation  
der Fakultät für Biologie  
der Ludwig-Maximilians-Universität München

vorgelegt von  
Christina Ulrike Bauer

München, den 12.04.2016

Erstgutachter: Prof. Dr. Heinrich Leonhardt

Zweitgutachter: Prof. Dr. Dirk Eick

Tag der Abgabe: 12.04.2016

Tag der mündlichen Prüfung: 14.10.2016



## Content

<b>Summary</b>	5
<b>Zusammenfassung</b>	7
<b>Aims of the thesis</b>	9
<b>1. Introduction</b>	10
1. 1 Epigenetics and DNA modifications	10
1.1.1 Genome organization and epigenetic regulation	10
1.1.2 DNA methylation	11
1.1.3 TET proteins and enzymatic DNA oxidation	13
1.2 Post-translational modifications of proteins	16
1.2.1 PTM dynamics	16
1.2.2 Histone modifications	18
1.2.3 Protein phosphorylation and 14-3-3 proteins	19
1.2.4 <i>O</i> -linked N-Acetylglucosamination ( <i>O</i> -GlcNAcylation) by the glycosyltransferase OGT	20
1.3 Early embryonic development	22
1.3.1 Early <i>in vivo</i> development of mouse embryos	22
1.3.2 <i>In vitro</i> models of differentiation	23
<b>2. Results</b>	25
2.1 Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives	25
2.2 TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation	49
2.3 Phosphorylation of TET proteins is regulated via <i>O</i> -GlcNAcylation by the <i>O</i> -linked <i>N</i> -Acetylglucosamine transferase (OGT)	70
2.4 A modular open platform for systematic functional studies under physiological conditions	87
2.5 The impact of DNA modifications on chromatin composition in embryonic stem cells and epiblast-like cells	121
<b>3. Discussion</b>	155
3.1 Regulation of TET proteins by interaction partners and PTMs	155
3.1.1 TET proteins, base excision repair, and DNA demethylation	155
3.1.2 Phosphorylation and <i>O</i> -GlcNAcylation of TET proteins	157
3.1.3 TET1 and chromatin modifiers	163
3.2 Interplay between cytosine modifications and chromatin proteins	166
<b>References</b>	173

## Content

<b>Annex</b>	183
Abbreviations	183
List of publications	185
Statutory declaration and statement	186
Acknowledgements	187

## Summary

Epigenetic processes ensure propagation of cell identity and gene expression patterns across cell divisions without affecting the primary DNA sequence. Chemical modifications of histones and DNA shape chromatin structure and are the most important mechanisms of epigenetic regulation. This thesis focuses on the interplay of DNA modifications with chromatin proteins and on the biological regulation and function of the DNA modifiers TET1, TET2, and TET3 in mammalian cells. Methylation of cytosine is one of the longest known epigenetic marks and is set by DNA methyltransferases (DNMTs). In 2009, the family of TET proteins was discovered to oxidize methylcytosine to hydroxymethylcytosine, formylcytosine, and carboxylcytosine. Since then, unraveling of the cellular function of these newly discovered cytosine variants and understanding the complex regulatory role of TET proteins has been a major goal in epigenetic research.

In this thesis, we identify and characterize several protein-protein-interaction partners of TET1, TET2, and TET3, including OGT, base excision repair proteins and a variety of chromatin modifiers. We show that TET proteins are associated with multiple DNA glycosylases and that the concerted action of TET proteins and the base excision repair machinery leads to reactivation of a previously silenced reporter gene in embryonic stem cells (ESCs). Furthermore, we discover that all three TET proteins are subjected to dynamic post-translational modifications (PTMs), in particular phosphorylation and *O*-GlcNAcylation. The *O*-GlcNAcylation mark is set by the enzyme OGT and increase of *O*-GlcNAcylation of TETs leads to a decrease in phosphorylation sites and occupancy. Moreover, we demonstrate that TET1 interacts with proteins involved in histone modification pathways such as methylation of H3K4 or deacetylation. Taken together, our findings suggest that TET proteins are dynamically regulated by interacting proteins and PTMs and strongly support the hypothesis that TETs contribute to the formation of both active and inactive chromatin states.

To understand the crosstalk between cytosine modifications and chromatin proteins in ESCs and during differentiation, we apply two complementary proteomics-based approaches. The first approach identifies direct reader proteins of cytosine variants by incubation of cellular lysates with differentially modified DNA baits and shows that these readers greatly differ in ESCs and cells of the neuronal lineage. The second approach allows us to study the impact of cytosine variants in their natural chromatin context. We characterize the total chromatin composition in cells that lack major DNA modifiers, DNMTs, TETs, or TDG, and therefore carry an altered DNA modification landscape. In DNMT-deficient cells, only canonical cytosine is present, whereas TET-deficient cells are depleted of all oxidative cytosine variants. In contrast, *Tdg*-knockout cells display an accumulation of formyl- and carboxylcytosine. We show that these perturbations of genomic DNA modifications greatly affect chromatin composition in ESCs, in particular that loss

## Summary

of DNMTs results in a depletion of transcription factors like KLF4, OCT4, and NonO, as well as of repressive chromatin marks, and that knockout (KO) of TDG results in recruitment of a variety of DNA repair factors to the chromatin of ESCs. Both approaches shed light onto the complex interplay between DNA modifications and proteinaceous chromatin composition and enhance the current knowledge on how cytosine variants are interpreted by cellular proteins and contribute to epigenetic regulation.

## Zusammenfassung

Die Weitergabe von Zellidentität und Genexpressionsmustern während der Zellteilung erfolgt über epigenetische Prozesse, ohne dass die DNA-Sequenz verändert wird. Chemische Modifikationen von Histonen und DNA beeinflussen die Chromatinstruktur und sind die wichtigsten epigenetischen Regulationsmechanismen. Der Fokus dieser Arbeit liegt auf dem Zusammenspiel von DNA-Modifikationen mit Chromatinproteinen und auf der biologischen Regulation und Funktion der DNA-modifizierenden Enzyme TET1, TET2, und TET3 in Säugetierzellen. Cytosinmethylierung ist eines der am längsten bekannten epigenetischen Signale und wird durch DNA-Methyltransferasen (DNMTs) katalysiert. Die TET-Proteinfamilie, die Methylcytosin zu Hydroxymethylcytosin, Formylcytosin, und Carboxylcytosin oxidiert, wurde erst im Jahr 2009 entdeckt. Seitdem ist das Verständnis der zellulären Funktion dieser neu beschriebenen Cytosinvarianten und die komplexe Rolle der TET-Proteine eine zentrale Frage der Epigenetik.

In dieser Arbeit identifizieren und charakterisieren wir mehrere Protein-Protein-Interaktionspartner von TET1, TET2, und TET3, darunter OGT, Proteine der Basenexzisionsreparatur, und weitere Chromatin-modifizierende Proteine. Wir zeigen, dass TET-Proteine mit mehreren DNA-Glykosylasen assoziiert sind und dass das koordinierte Zusammenspiel von TET-Proteinen mit der Basenexzisionsreparaturmaschinerie zur Reaktivierung eines zuvor stillgelegten Reportergens in embryonalen Stammzellen (ES-Zellen) führt. Wir beschreiben weiterhin, dass alle drei TET-Proteine post-translational modifiziert werden, genauer gesagt phosphoryliert und *O*-GlcNAcyliert. Das Enzym OGT katalysiert diese *O*-GlcNAcylierung und ein Anstieg derselben führt zu einer Reduktion von Phosphorylierungsstellen und Phosphorylierungsstärke. Außerdem zeigen wir, dass TET1 mit Proteinen interagiert, die eine Rolle bei der Modifikation von Histonen spielen, z.B. Methylierung von H3K4 oder Deacetylierung. Zusammenfassend legen unsere Daten nahe, dass TET-Proteine durch Interaktionspartner und post-translationale Modifikationen dynamisch reguliert werden und stützen die Hypothese, dass TETs sowohl zur Bildung von aktiven als auch inaktiven Chromatinkonformationen beitragen.

Um das Zusammenspiel von Cytosinmodifikationen und Chromatinproteinen in ES-Zellen und während der Zelldifferenzierung zu verstehen, wenden wir zwei komplementäre, Proteomik-basierte Ansätze an. Mit dem ersten Ansatz werden direkte "Lese proteine" der Cytosinvarianten identifiziert, indem verschieden modifizierte DNA mit Zelllysat inkubiert wird. Dieser Ansatz zeigt, dass in ES-Zellen und neuronalen Zellen unterschiedliche Lese proteine an die DNA-Modifikationen binden. Der zweite Ansatz ermöglicht es, den Einfluss der Cytosinvarianten in ihrer natürlichen Chromatinumgebung zu untersuchen. Dazu charakterisieren wir die

Chromatinzusammensetzung in Zellen, denen zentrale DNA-modifizierende Enzyme, DNMTs, TETs oder TDG, fehlen und die deshalb eine veränderte Zusammensetzung von DNA-Modifikationen zeigen. Zellen ohne DNMTs haben nur kanonisches Cytosin, während Zellen ohne TETs keine oxidativen Cytosinvarianten haben. Im Gegensatz dazu zeigen Zellen ohne TDG eine Anreicherung von Formyl- und Carboxylcytosin. Wir zeigen, dass diese Störungen der genomischen DNA-Modifikationsmuster einen großen Einfluss auf die Chromatinzusammensetzung in ES-Zellen haben und dass der Verlust von DNMTs zu einer Verminderung sowohl von Transkriptionsfaktoren wie KLF4, OCT4 und NonO, als auch von inaktivierenden Chromatinkomponenten führt. Des Weiteren werden zahlreiche DNA-Reparaturproteine in Zellen, denen TDG fehlt, zum Chromatin rekrutiert. Beide Ansätze tragen zum Verständnis des komplexen Zusammenspiels zwischen DNA-Modifikationen und Chromatinzusammensetzung bei und erweitern das gegenwärtige Wissen darüber, wie Cytosinvarianten durch zelluläre Proteine interpretiert werden und wie sie zur epigenetischen Regulation beitragen.

## Aims of the thesis

DNA modifications, in particular cytosine modifications, are a crucial part of epigenetic regulatory mechanisms in mammals. There are four described chemical variants of cytosine: methylcytosine, hydroxymethylcytosine, formylcytosine, and carboxylcytosine. The first mark, methylcytosine, is set by DNMTs. The latter three are generated by the family of TET dioxygenases through consecutive oxidation of methylcytosine. The goal of this dissertation is to characterize the cellular regulation of TET proteins, and the interplay of cytosine modifications with gene expression and chromatin-associated proteins in mammalian cells. The following aims are addressed in particular:

First, we wanted to identify reader proteins of the different cytosine modifications and analyze if and how these readers change over the course of neuronal differentiation. To this end, we used differentially modified DNA as a bait to pull down binding proteins from cellular lysates of ESCs and neuronal cells. Selected readers were characterized *in vitro* to gain insights into their biological function.

Second, we asked how the oxidized cytosine variants influence gene expression and which cellular pathways contribute to this transcriptional regulation. We therefore applied a reporter gene assay and studied the impact of DNA modifications on the expression of the reporter gene dependent on the availability of different DNA glycosylases.

Third, we aimed to understand regulation of TET proteins by PTMs. Tandem mass spectrometry was performed to identify and quantify TET protein modification sites. We furthermore investigated the dynamic interplay between different PTMs and their distribution across the TET primary sequence.

Fourth, we screened for proteins within the regulatory environment of TET1 by proximity-dependent labeling. We wanted to identify chromatin modifiers that contribute to TET1 function and regulation.

Lastly, we aimed to dissect the complex crosstalk between DNA modifications and chromatin-associated proteins. To this end, we performed chromatin enrichments in *Dnmt*-triple-KO, *Tet*-triple-KO, and *Tdg*-KO cells and investigated how the altered DNA modification patterns of these KO cell lines influence proteinaceous chromatin composition.

## 1. Introduction

### 1. 1 Epigenetics and DNA modifications

#### 1.1.1 Genome organization and epigenetic regulation

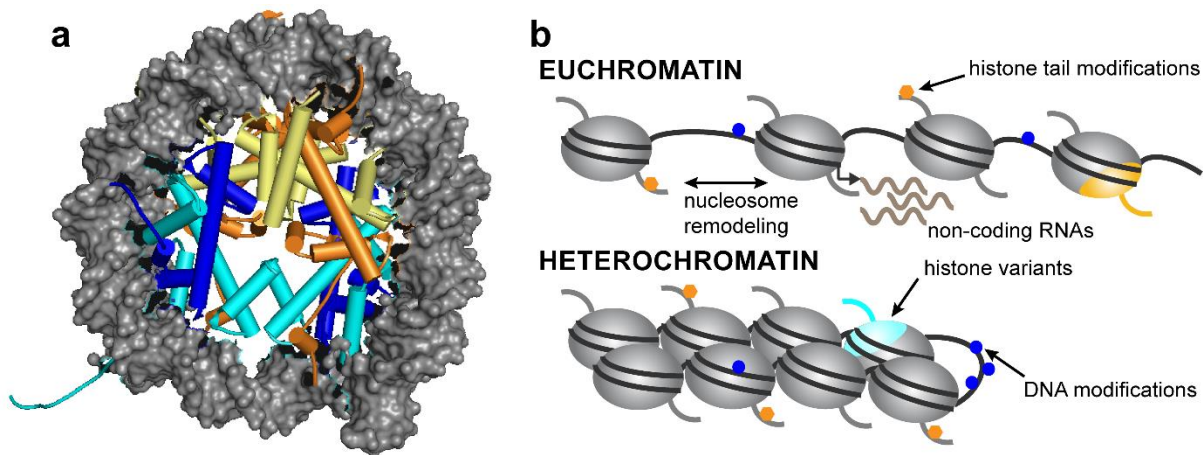
Ever since Oswald Avery discovered that DNA is the molecule that contains the genetic information of an organism (Avery *et al.*, 1944), sequence, interpretation, organization, replication, and structure of this genetic information has been of great interest to scientists and the general public. Today, we know that the human haploid genome consists of 3 billion base pairs (bp) (Morton, 1991), encoding approximately 20 thousand genes (Ewing and Green, 2000; Venter *et al.*, 2001; Pertea and Salzberg, 2010). Despite their various shapes and functions, almost all cells in the human body contain this very same set of genes and they all originate from one single cell, the fertilized egg (called zygote). As a term that characterizes unfolding of the developmental program based on the genome of the zygote, the word “epigenetic” was first used by C.H. Waddington in his book “Introduction to Modern Genetics” to describe a process that is predetermined in the genome, but relies on properties beyond the genetic code (Waddington, 1939). Nowadays, epigenetics is defined more specifically as “changes in gene expression which occur in organisms with differentiated cells, and the mitotic inheritance of given patterns of gene expression [...] not based on changes in DNA sequence” (Holliday, 2014).

Genome organization and epigenetic regulation of gene expression are closely linked. The smallest structural unit of the genome is the nucleosome (Kornberg, 1974; Olins and Olins, 1974). It consists of about 147 bp of DNA wrapped around a protein octamer which is formed by two copies of each of the four core histones, H2A, H2B, H3, and H4 (Luger *et al.*, 1997). The flexible N-terminal histone tails stick out of the core particle and are subjected to PTMs and protein binding (Figure 1a). The complex of genomic DNA with proteins such as histones, transcription factors, repressors, and remodelers is generally referred to as “chromatin” (Arney and Fisher, 2004), a term introduced by W. Flemming around 1880 to describe the easily stained substance within the cell nucleus (Olins and Olins, 2003). The packaging of nucleosomes into loose or dense structures defines eu- and heterochromatin, two states which are associated with transcriptional activation or silencing, respectively (Arney and Fisher, 2004) (Figure 1b).

Heterochromatin is categorized into two groups: constitutive and facultative heterochromatin. The first contains permanently inactivated genomic regions such as repetitive sequences and gene-poor areas. The latter describes chromatin that is silenced in a cell-type dependent manner. Both the constitutive and the facultative heterochromatin state is preserved during mitotic cell division (Richards and Elgin, 2002).



Epigenetic control of chromatin states is achieved by several interdependent mechanisms, including PTMs of histones, in particular of histone tails, chemical modification of DNA, chromatin remodeling, incorporation of histone variants, and non-coding RNA (Figure 1b). These mechanisms influence chromatin assembly and accessibility and thereby contribute to transcriptional regulation (Wolffe and Guschin, 2000; Grewal and Moazed, 2003; Arney and Fisher, 2004; Holliday, 2014).



**Figure 1: Mechanisms of genome organization**

a) Structure of the nucleosome (PDB accession number 1AOI, (Luger *et al.*, 1997)). DNA is depicted as gray surface model, histones are schematically shown with alpha helices as cylinders. H2A and H2B, as well as H3 and H4, form a heterodimer, respectively. Dark blue: histone H4, cyan: histone H3, orange: histone H2A, yellow: histone H2B.

b) Schematic illustration of epigenetic mechanisms.

### 1.1.2 DNA methylation

DNA methylation is one of the most widely studied and best understood epigenetic marks. In mammals, it occurs predominantly at the carbon 5 atom of cytosine in a CG dinucleotide context (Smith and Meissner, 2013). In general, methylcytosine (mC) is associated with transcriptional repression, and the majority is found at repetitive elements or transposons (Goll and Bestor, 2005). Three enzymes, called DNA methyltransferases (DNMTs), are responsible for methylation of DNA: DNMT1, DNMT3A, and DNMT3B (Bestor *et al.*, 1988; Okano *et al.*, 1998; Rottach *et al.*, 2009). While the catalytic methyltransferase domain is highly conserved in these three proteins, DNMT1 additionally harbors a large regulatory N-terminus containing several subdomains (Figure 2a). Both DNMT3s contain an ADD (ATRX, DNMT3, DNMT3L) domain that consists of two zinc fingers, packed together by hydrophobic interactions (Argentaro *et al.*, 2007). The PWWP domain of DNMT3A and DNMT3B likely functions as a protein-protein interaction motif (Stec *et al.*, 2000) and is required for association with pericentric heterochromatin (Chen *et al.*, 2004). The enzymatically inactive associated factor DNMT3L plays a central role in germ cell

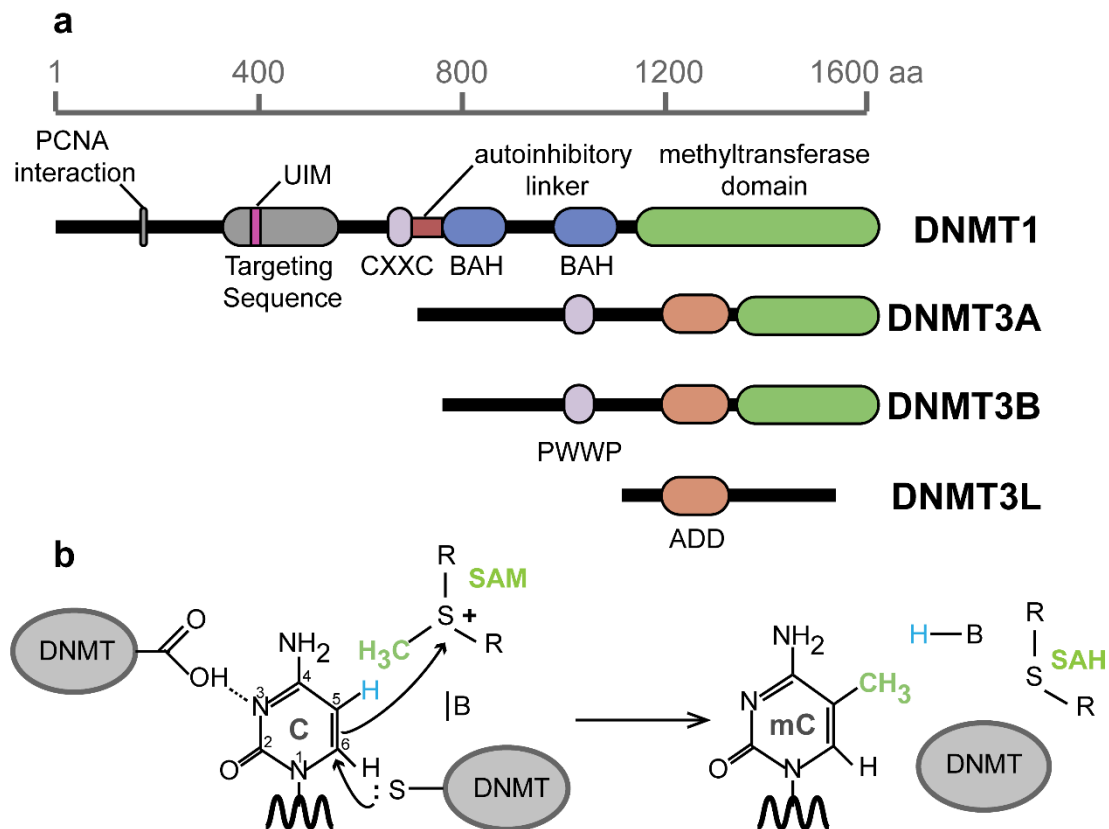
## 1. Introduction

development, establishment of maternal methylation imprints and methylation of retrotransposons in sperm (Bourc'his *et al.*, 2001; Goll and Bestor, 2005). DNMT1 has a highly structured N-terminus with a domain required for interaction with the replication clamp PCNA (proliferating cell nuclear antigen) and a sequence responsible for targeting to replication foci that contains a UIM (ubiquitin interaction motif) (Qin *et al.*, 2015). A CXXC-type zinc finger is connected to one of two BAH (bromo-adjacent homology) domains by an autoinhibitory linker (Song *et al.*, 2011). BAH domains are often found in chromatin proteins and are connected to transcriptional repression (Callebaut *et al.*, 1999).

The three catalytically active DNMTs all share a common reaction mechanism, requiring S-Adenosylmethionine (SAM) as a methyl-donor. A covalent bond is formed between the SH-group of a conserved cysteine and carbon 6 of the base. This nucleophilic attack is assisted by a glutamate residue of the ENV motif, transiently protonating nitrogen 3. After transfer of the methyl group, the proton on C5 is bound by an unknown basic residue, allowing re-formation of the C5-C6 double bond. S-Adenosyl-homocysteine (SAH) is released as a side product of the reaction (Santi *et al.*, 1983; Chen *et al.*, 1991; Bestor, 2000; Goll and Bestor, 2005) (Figure 2b). However, formation of the covalent bond between the cysteine and the base seems not to be strictly required for catalytic activity, whereas mutation of glutamate in the ENV motif completely abolishes methyltransferase activity (Reither *et al.*, 2003).

Within the cell, the different DNMTs have distinct roles. DNMT1 has been termed the "maintenance" methyltransferase, because it is preferentially active on hemi-methylated DNA, ensuring inheritance of the methylation mark during DNA replication (Hermann *et al.*, 2004). *In vivo*, DNMT1 activity is tightly regulated. The protein associates with the replication machinery via a targeting sequence at the N-terminus (Leonhardt *et al.*, 1992) and directly interacts with PCNA (Chuang *et al.*, 1997). This interaction enhances methylation efficiency, but is not essential for catalytic activity (Schermelleh *et al.*, 2007). DNMT1 furthermore interacts with UHRF1 (Ubiquitin-like PHD and RING finger domain-containing protein 1), a protein with a preference for hemi-methylated DNA that is required for maintenance methylation (Bostick *et al.*, 2007). UHRF1 also directly influences DNMT1 stability by ubiquitination, thereby triggering proteasomal degradation (Du *et al.*, 2010; Qin *et al.*, 2011).

DNMT3A and DNMT3B are the so-called *de novo* methyltransferases, which are responsible for establishing DNA methylation patterns during early embryonic development (Okano *et al.*, 1999). Both enzymes act redundantly in methylation of the bulk genome e.g. at repetitive elements, but also have distinct roles. DNMT3B, for example, is responsible for methylation of CpG islands and gene bodies during embryonic implantation (Auclair *et al.*, 2014). DNMT3A has been shown to have an important role in germ cell development, likely in cooperation with DNMT3L (Chedin *et al.*, 2002; Kaneda *et al.*, 2004; Goll and Bestor, 2005).



**Figure 2: DNA methylation in mammals**

a) Schematic depiction of the domain architecture of murine DNMT proteins (Uniprot-IDs P13864, O88508, O88509, Q9CWR8). The catalytically active DNMTs share a C-terminal methyltransferase domain that carries out the enzymatic reaction.

b) Reaction mechanism of DNMTs. The methyl group of the co-factor SAM (S-Adenosylmethionine) is transferred to C5 of cytosine, accompanied by formation of a covalent complex between the enzyme and the base (Goll and Bestor, 2005).

Most of mammalian DNA methylation occurs at repetitive elements such as pericentromeric repeats, long interspersed nuclear elements (LINEs) or long terminal repeats (LTRs) (Smith and Meissner, 2013). Methylation of gene promoters is generally associated with transcriptional repression, the exact read-out mechanism for the methylation mark still being elusive (Baubec and Schubeler, 2014). Promoters with high CG content, termed CpG islands (CGIs), are usually not methylated, highly active, and often associated with housekeeping genes (Meissner *et al.*, 2008; Deaton and Bird, 2011).

### 1.1.3 TET proteins and enzymatic DNA oxidation

While DNA methylation has been intensively investigated for several decades, demethylation, i.e. removal of the methyl group, has long remained a mystery. Generally, there are two different principles of DNA demethylation. First, passive dilution of the methyl mark through several cycles

## 1. Introduction

of DNA replication and second, active removal of the methyl group. Passive demethylation can be achieved by retention of DNMT1 in the cytoplasm which has been described for pre-implantation embryos (Cardoso and Leonhardt, 1999). The principles of active demethylation, however, have been discovered only recently.

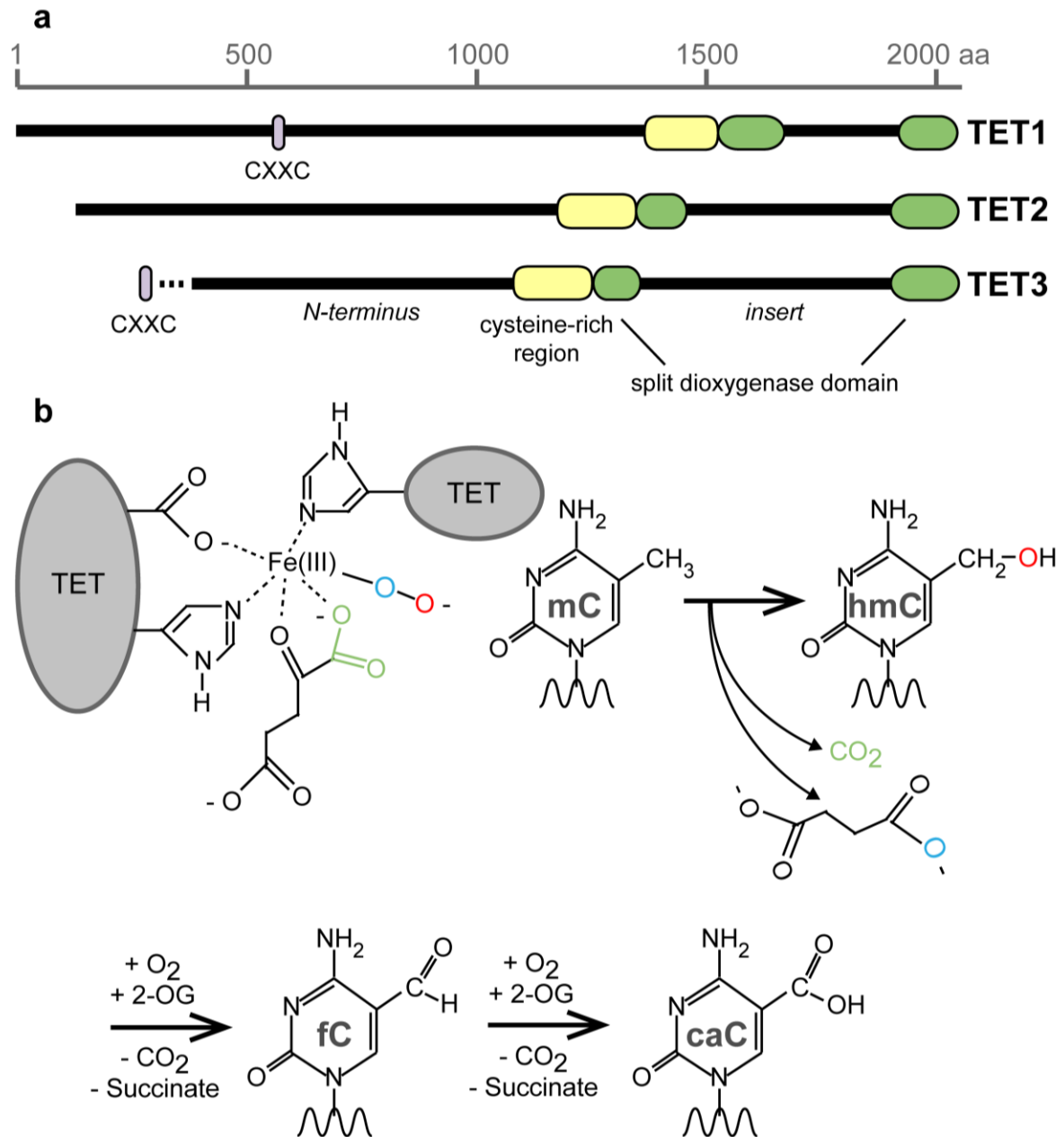
In 2009, a family of proteins, the TET proteins, has been described to oxidize mC to hydroxymethylcytosine (hmC) (Iyer *et al.*, 2009; Tahiliani *et al.*, 2009), which was immediately proposed to be an intermediate step of active DNA demethylation. Two years later, TET proteins have been shown to further oxidize hmC to formylcytosine (fC) and carboxylcytosine (caC). The higher oxidized cytosine variants, fC and caC, can then be removed by the Thymine-DNA glycosylase (TDG) (He *et al.*, 2011; Maiti and Drohat, 2011). The name “TET proteins” was first introduced by Lorschbach *et al* who described and cloned *TET1* as a fusion partner of the *MLL* gene in acute myeloid leukemia (AML) in a t(10;11)q(22;23) translocation event (TET: ten-eleven-translocation) (Lorschbach *et al.*, 2003).

In human and mouse, there are three protein family members, TET1, TET2, and TET3, that can carry out the same enzymatic reactions (Lorschbach *et al.*, 2003; Iyer *et al.*, 2009). All three TET proteins contain a C-terminal catalytic domain, consisting of two parts separated by a presumably unstructured insert and preceded by a cysteine-rich region (Iyer *et al.*, 2009) (Figure 3a). TET proteins belong to the family of Fe(II) and 2-oxoglutarate (2-OG, also  $\alpha$ -ketoglutarate) dependent dioxygenases, which oxidize their target structure while decarboxylating the co-factor 2-OG to succinate (Loenarz and Schofield, 2009) (Figure 3b).

The crystal structure of human TET2 revealed that the two halves of the dioxygenase domain are closely packed against the cysteine-rich region to form the catalytic core of the enzyme. In addition to the indispensable Fe(II), three zinc ions are coordinated by the protein to stabilize DNA binding and enhance catalytic efficiency. TET2 displays a preference for mCG dinucleotide sequences and flips mC out of the DNA double strand into its catalytic pocket to carry out the oxidation reaction. The catalytic pocket provides sufficient space to also allow the subsequent oxidation steps from hmC to fC and caC (Hu, Li, *et al.*, 2013).

In addition to the well-understood and conserved catalytic domain, TET proteins harbor two large amino acid stretches that are of low complexity and presumably unstructured: The N-terminus and the insert region between the two parts of the catalytic domain (Iyer *et al.*, 2009). The function of these parts of the proteins is largely unknown. They are not required for catalytic activity and in case of the N-terminal sequence of TET1, but not TET2, seem to decrease catalytic efficiency (Tahiliani *et al.*, 2009; Hu, Li, *et al.*, 2013). Human TET1 and TET3 harbor a CXXC-type zinc finger at the N-terminus, whereas TET2 has no zinc finger in both human and mouse. From an evolutionary perspective, TET2 likely carried a zinc finger module that underwent gene inversion. This zinc finger is now expressed as a separate protein, IDAX or CXXC4, which directly

binds TET2 and results in its downregulation via caspase activation (Ko *et al.*, 2013). Murine TET3 is expressed in two isoforms in neuronal stem cells, one expressed from an alternative upstream start codon that carries a CXXC zinc finger and one expressed from a downstream ATG that has no zinc finger (Iyer *et al.*, 2009; Liu *et al.*, 2013).



**Figure 3: Murine TET proteins**

a) Schematic representation of murine TET proteins and their domains (Uniprot-IDs Q3ERK3, Q4JK59, Q8BG87). The cysteine-rich region and the two halves of the dioxygenase domain are conserved for all three TET proteins, whereas the N-terminus and the insert region differ in sequence and size. TET1 contains an N-terminal CXXC-type zinc finger and murine TET3 has two isoforms, one with and one without zinc finger (Liu *et al.*, 2013).

b) Reaction mechanism of Fe(II) and 2-OG dependent dioxygenases. The catalytic iron forms a complex with two histidine and one aspartate residue of the enzyme, with oxygen and the co-factor 2-OG. One oxygen atom is transferred to the base to convert mC to hmC while releasing succinate

## 1. Introduction

and carbon dioxide. Similarly, the subsequent oxidation steps to fC and caC are performed (Muller *et al.*, 2005; Loenarz and Schofield, 2009).

Oxidation of mC by TET proteins has been suggested as a potential mechanism for active DNA demethylation and indeed, shortly after its initial characterization, TET1 has been shown to induce demethylation in brain tissue. The then proposed mechanism involves coupling of hmC formation to deamination by AID/APOBEC family proteins and subsequent removal of the generated hydroxymethyluracil (hmU) by base excision repair (BER) pathways (Guo *et al.*, 2011). Shortly after, TET proteins have been discovered to also catalyze the oxidation reactions to fC and caC, which can both be directly excised by TDG and replaced with unmodified C by BER (He *et al.*, 2011; Maiti and Drohat, 2011). Furthermore, AID/APOBEC deaminases disfavor binding of C with bulky adducts (Nabel *et al.*, 2012), raising questions about the biological relevance of oxidation-deamination-coupled DNA demethylation. In 2014, catalytic activity of TET proteins on thymine was reported, directly generating hmU without a deamination step and thus providing an explanation for the previously reported genomic hmU levels (Pfaffeneder *et al.*, 2014). The cellular role of hmU, however, is still elusive.

In summary, the regulation and function of TET proteins is currently intensively investigated. Many open questions remain about the role of the oxidized base derivatives in the context of DNA demethylation and beyond.

## 1.2 Post-translational modifications of proteins

### 1.2.1 PTM dynamics

Modifications of DNA bases and of the N-terminal histone tails constitute the two major epigenetic mechanisms. Not only histones, but almost any protein can be chemically modified at amino acid side chains, a process that is generally referred to as post-translational modification (PTM). These PTMs have important functions in protein activity, stability, and localization, in protein-protein interactions, and in a variety of cellular processes such as cell cycle, DNA damage response and signaling cascades (Karve and Cheema, 2011). The dynamics of PTMs are controlled by enzymatic addition or removal of the modification by specific “writer” or “eraser” domains, respectively (Beltrao *et al.*, 2013). Presence or absence of a modification is often coupled to cellular function through recognition by a “reader” protein that can specifically bind to the modified amino acid (Seet *et al.*, 2006; Beltrao *et al.*, 2013). More than 400 different types of PTMs have been described to date (Minguez *et al.*, 2012). Table 1 provides an overview about the most common ones and Figure 4 depicts the chemical structure of selected PTMs.

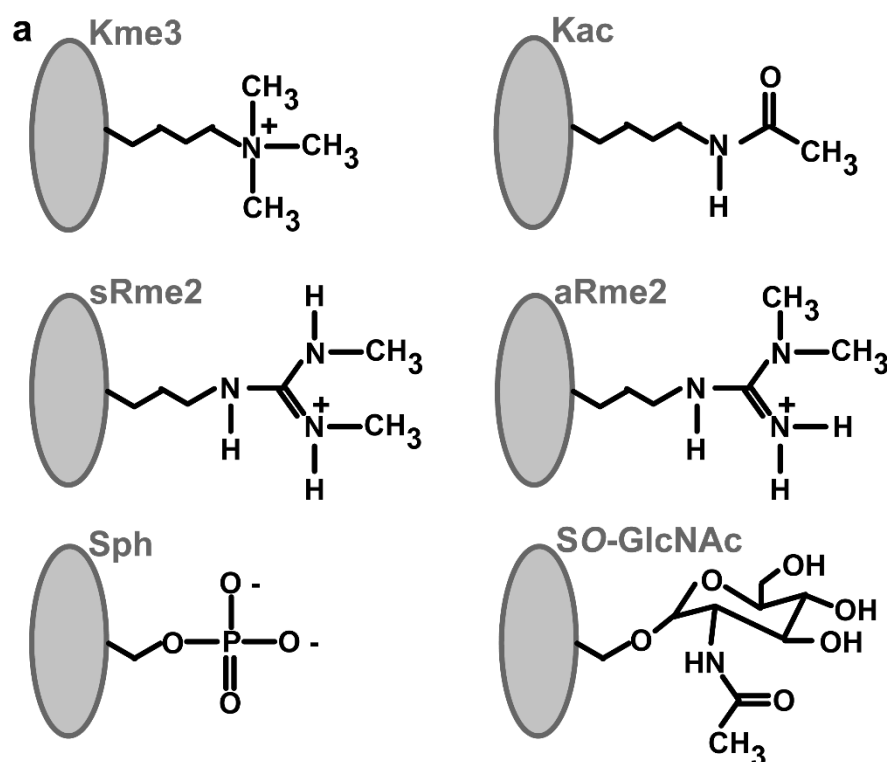
Modifications such as phosphorylation, acetylation and glycosylation have been detected in almost every living organism, suggesting a long evolutionary history (Beltrao *et al.*, 2013). Furthermore, many PTMs are found in combinations or patterns that can be associated with

distinct cellular states. Among the best-studied examples are histone tails and the CTD (C-terminal domain) of RNA polymerase II that are subjected to a large number of different combinatorial PTMs (Lothrop *et al.*, 2013).

modification	amino acid
acetylation	lysine
deamination	asparagine, glutamine
glycosylation	serine, threonine (O-linked) asparagine (N-linked)
hydroxylation	proline
methylation	lysine (mono-, di-, trimethylation) arginine (monomethylation, symmetric or asymmetric dimethylation)
oxidation	cysteine
phosphorylation	serine, threonine, tyrosine
sumoylation	lysine
ubiquitination	lysine

**Table 1: Common protein PTMs**

Summary of common protein post-translational modifications. Sumoylation and ubiquitination are not small chemical modifications, but (poly)peptides, which are covalently coupled to the target protein (Seo and Lee, 2004; Bedford, 2007).



**Figure 4: Chemical structures of selected protein PTMs**

Depicted are the following amino acid side chain modifications: lysine trimethylation (Kme3), lysine acetylation (Kac), symmetric and asymmetric arginine methylation (sRme2 and aRme2, respectively), serine phosphorylation (Sph), and serine O-GlcNAcylation (SO-GlcNAc) (Zhang and Reinberg, 2001; Bedford, 2007; Patel *et al.*, 2011; Endicott *et al.*, 2012; Harwood and Hanover, 2014).

## 1. Introduction

### 1.2.2 Histone modifications

In epigenetics, PTMs of histone tails are of particular interest. The complex patterns of modifications that can be found at histone tails have also been referred to as the “histone code”, which is considered to be an expansion of the genetic code (Jenuwein and Allis, 2001). Acetylation, methylation, phosphorylation, and ubiquitination are the most important PTMs on histone tails that have been correlated with differential transcriptional states or cell cycle stages (Bhaumik *et al.*, 2007).

Histone acetylation plays a crucial role in chromatin assembly and transcriptional control. Histone proteins are strongly acetylated after synthesis and deacetylated following incorporation into freshly replicated DNA. Since acetylation neutralizes the positive charge of lysine residues and thereby weakens the histone-DNA-association, histone acetylation is generally more associated with euchromatin. Enzymes responsible for acetylation and deacetylation are histone acetyltransferases (HATs) and histone deacetylases (HDACs), respectively (Shahbazian and Grunstein, 2007).

Several specific modification sites have been studied in great detail over the last years. For example, trimethylation of histone 3 at lysine 27 (H3K27me3) silences gene expression (Cao and Zhang, 2004). This mark is set by the polycomb repressive complex 2 (PRC2), a protein complex that has been described first in the fruit fly *Drosophila melanogaster* (Schuettengruber *et al.*, 2007). In human or mouse, PRC2 consists of five components: EZH2 (enhancer of zeste homolog 2), EED, SUZ12 (suppressor of zeste 12 protein homolog), and the histone binding proteins RBBP4 and RBBP7. The isolated complex possesses methyltransferase activity towards H3K9 and H3K27, the latter being the primary target. H3K27me3, but not H3K9me3, can in turn be bound by Polycomb Repressive Complex 1 (PRC1) (Kuzmichev *et al.*, 2002).

PRC1 contains the subunit RING1B and catalyzes ubiquitination of H2AK119 (Bhaumik *et al.*, 2007). KO of the *Ring1b* gene in mouse is embryonic lethal at the gastrulation stage (Voncken *et al.*, 2003). However, the ubiquitination activity seems to be dispensable for gene silencing and embryonic development (Eskeland *et al.*, 2010; Illingworth *et al.*, 2015), leaving the importance and function of H2AK119ub unclear.

Other well studied repressive histone marks are H3K9me3 and H4K20me3. These modifications are predominantly found at repetitive elements, centromeres, and telomeres, contributing to stable long-term silencing of constitutive heterochromatin (Martens *et al.*, 2005; Mikkelsen *et al.*, 2007).

Phosphorylation of histone H3 at serine 10 (H3S10ph) has long been known to induce chromosome condensation during mitosis and meiosis (Gurley *et al.*, 1978; Nowak and Corces, 2004). H3S10ph during mitosis is regulated by the opposing activities of AURORA-B kinases and



PP1 phosphatases (Nowak and Corces, 2004; Fu *et al.*, 2007). Furthermore, H3S10ph has been implicated in the rapid transcriptional changes upon response to external stimuli such as EGF (epidermal growth factor), FGF2 (fibroblast growth factor 2), TNF- $\alpha$  (tumor necrosis factor  $\alpha$ ), or stress signaling (Nowak and Corces, 2004).

Trimethylation of lysine 4 of histone H3 (H3K4me3) is an active histone mark that can be found at the promoter regions of transcribed genes. This modification is set by trithorax group (trxG) proteins, more specifically by trxG SET domain factors. In humans, this function is carried out by MLL complexes. Other trxG proteins are not involved in H3K4 methylation, but are ATP-dependent chromatin remodelers like the SWI/SNF complex (Schuettengruber *et al.*, 2007). Interestingly, some gene promoters carry both the activating H3K4me3 and the repressive H3K27me3 mark. This phenomenon has been described for silent, lineage-specific genes in ESCs and the respective promoters have been termed “bivalent” (Azuara *et al.*, 2006; Mikkelsen *et al.*, 2007). In ESCs, active (H3K4me3-marked) and bivalent (H3K4me3/H3K27me3-marked) promoters usually have a high CG density and low levels of mC, whereas inactive promoters with intermediate or low CG frequency tend to be methylated (Fouse *et al.*, 2008). Thus, the signature and combinations of epigenetic marks at promoter regions provide information about the transcriptional status of the gene.

### 1.2.3 Protein phosphorylation and 14-3-3 proteins

Among all PTMs, phosphorylation is the most widely studied and best understood. In mouse, more than half of all proteins are phosphorylated at varying sites (Huttlin *et al.*, 2010). Phosphorylation can have several implications on protein function, e.g. enzyme activation or inhibition, recruitment or repulsion of co-factors, or changes in protein folding (Endicott *et al.*, 2012). Phosphorylation is enzymatically carried out by kinases; currently, the UniProt database lists 535 proteins with kinase domains in human and mouse each (UniProt, 2015). Kinases transfer a phosphate group from ATP to the protein substrate, the acceptor residue is either serine, threonine, or tyrosine. The number of target proteins per kinase varies from one to several hundred (Ubersax and Ferrell, 2007).

Phosphorylation introduces a bulky negatively charged adduct to the protein, thereby directly influencing intra- and intermolecular ion pairs and hydrogen bonds (Taverna *et al.*, 2007). On the one hand, this can induce a conformational change of the target protein, a mechanism often found in the allosteric activation of kinases. On the other hand, the modification can be recognized by reader modules of other proteins. Reader domains for phosphorylation are e.g. 14-3-3 domains, BRCT domains, TPR repeats, Polo boxes, FF domains, SRI domains, and WD40 domains (Seet *et al.*, 2006). The first two of these domains have also been implicated in recognition of phosphoserine residues on histones (Taverna *et al.*, 2007).

## 1. Introduction

There are seven paralogs of 14-3-3 proteins in mammals with the ability to homo- or heterodimerize (Taverna *et al.*, 2007). They are important components of phosphorylation-dependent signaling cascades and can mask localization sequences, prevent protein-protein interactions, and alter the accessibility of their target proteins. Close to 750 phosphorylated targets are bound by 14-3-3 proteins, many of which contain disordered sequences (Uhart and Bustos, 2014). Importantly, three 14-3-3 proteins have been described as interaction partners of TET2, namely YWHAH (14-3-3 subtype  $\eta$ ), YWHAG (subtype  $\gamma$ ), and YWHAQ (subtype  $\theta$ ) (Deplus *et al.*, 2013).

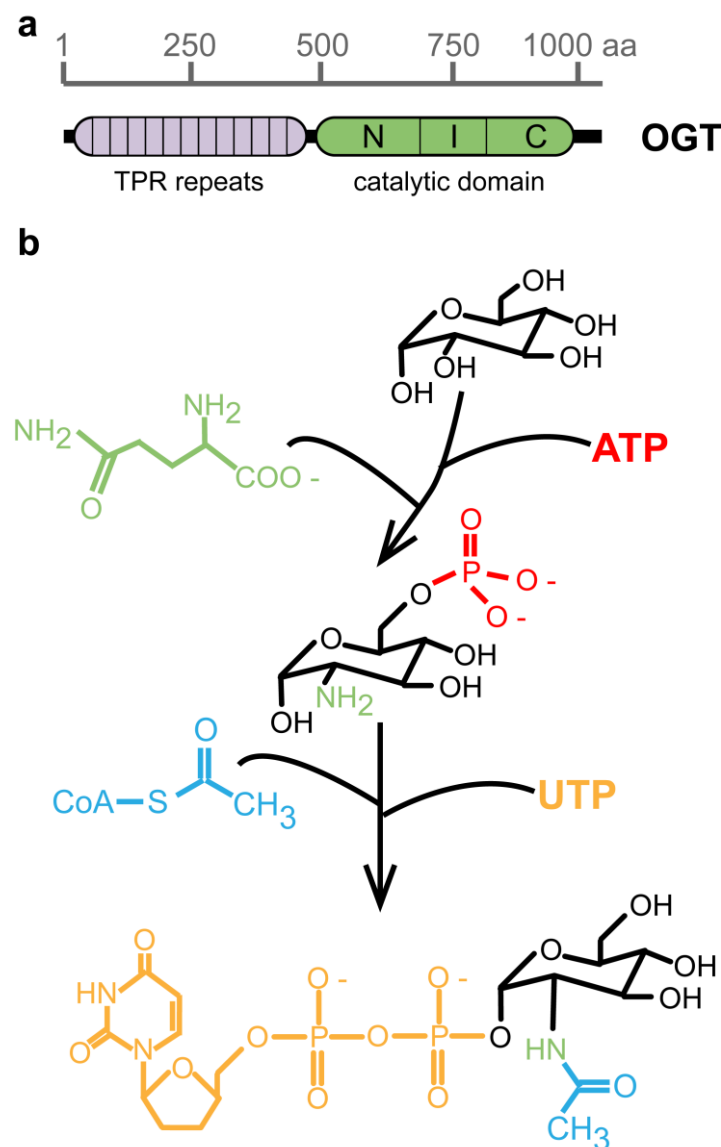
### **1.2.4 O-linked N-Acetylglucosamination (O-GlcNAcylation) by the glycosyltransferase OGT**

Kinases are among the largest families of enzymes encoded in the human genome. In contrast, O-GlcNAcylation is regulated by a single pair of proteins with opposing functions. The O-linked N-Acetylglucosaminetransferase (OGT) attaches the O-GlcNAc moiety to serine or threonine residues from where it can be removed by O-GlcNAcase (OGA) (Lubas *et al.*, 1997; Gao *et al.*, 2001; Harwood and Hanover, 2014). In human and mouse, the *OGT* gene is located on the X chromosome and is not only essential for embryonic development, but also for stem cell viability (Shafi *et al.*, 2000). The OGT protein consists of two major parts, an N-terminal series of TPRs (tetratricopeptide repeats) and the C-terminal catalytic domain (Figure 5a). OGT can be spliced into three isoforms that vary in the number of TPRs and thereby length of the N-terminus. These isoforms are the nucleocytoplasmic, the mitochondrial, and the short isoform, containing 12.5, 9.5, and 2.5 TPRs, respectively (Hanover *et al.*, 2003; Harwood and Hanover, 2014).

The substrate for OGT and donor of the O-GlcNAc moiety is UDP-GlcNAc (uridine diphosphate N-Acetylglucosamine), a product of the hexosamine biosynthetic pathway. This pathway senses the nutrient availability in the cell and is dependent on the cellular levels of glutamine, glucose, ATP, UTP, and Acetyl-CoA (Figure 5b) (Harwood and Hanover, 2014). Interestingly, O-GlcNAcylation is the only currently known type of glycosylation that occurs in the cytoplasm and nucleus of mammalian cells (Gambetta and Muller, 2015).

OGT has been described as a stable interaction partner of all three TET proteins by several groups (Chen *et al.*, 2012; Deplus *et al.*, 2013; Vella *et al.*, 2013; Ito *et al.*, 2014; Zhang *et al.*, 2014). In case of TET2, this interaction is mediated by the catalytic domain of TET and TPR 5 and 6 of OGT, but does not influence TET-dependent hmC generation. However, TET2 recruits OGT to chromatin and stimulates O-GlcNAcylation of Ser112 of H2B. All three components, OGT, TET2, and H2BS112O-GlcNAc, accumulate at transcription start sites (TSS) and seem to have an activating impact on gene expression (Chen *et al.*, 2012). Deplus *et al.* have described general stimulation of O-GlcNAcylation activity by TET2 and TET3, leading to glycosylation of HCF1 (host cell factor 1),

a component of the SET1/COMPASS complex that methylates H3K4 and is associated with transcriptionally active genes. Interestingly, TET2/OGT genomic binding sites are not enriched in mC or hmC (Deplus *et al.*, 2013). Expanding these results, Vella *et al.* have shown co-localization of TET1 and OGT at TSS, and locus-specific regulation of TET1 activity by OGT. They therefore suggest a dual role for TET1/OGT in transcriptional activation and repression (Vella *et al.*, 2013). Finally, OGT-mediated *O*-GlcNAcylation of TET3 has been demonstrated to lead to nuclear export of TET3. A similar effect can be observed when cultivating cells under high glucose conditions, directly linking TET protein regulation to metabolism (Zhang *et al.*, 2014).



**Figure 5: OGT and *O*-GlcNAcylation**

a) Schematic depiction of murine OGT, nucleocytoplasmic isoform with 12.5 TPRs (Uniprot-ID Q8CGY8). The most C-terminal TPR is truncated in all isoforms. The catalytic domain folds in three parts: the N-terminal (N), intervening (I), and C-terminal (C) domain. The glycosyl-donor UDP-GlcNAc is bound by the C-terminal part of the catalytic domain (Lazarus *et al.*, 2011).

## 1. Introduction

b) The hexosamine biosynthetic pathway. Several metabolites are used to modify glucose leading to the end product UDP-GlcNAc: ATP (red), glutamine (green), acetyl-CoA (blue), and UTP (orange) (Harwood and Hanover, 2014).

### 1.3 Early embryonic development

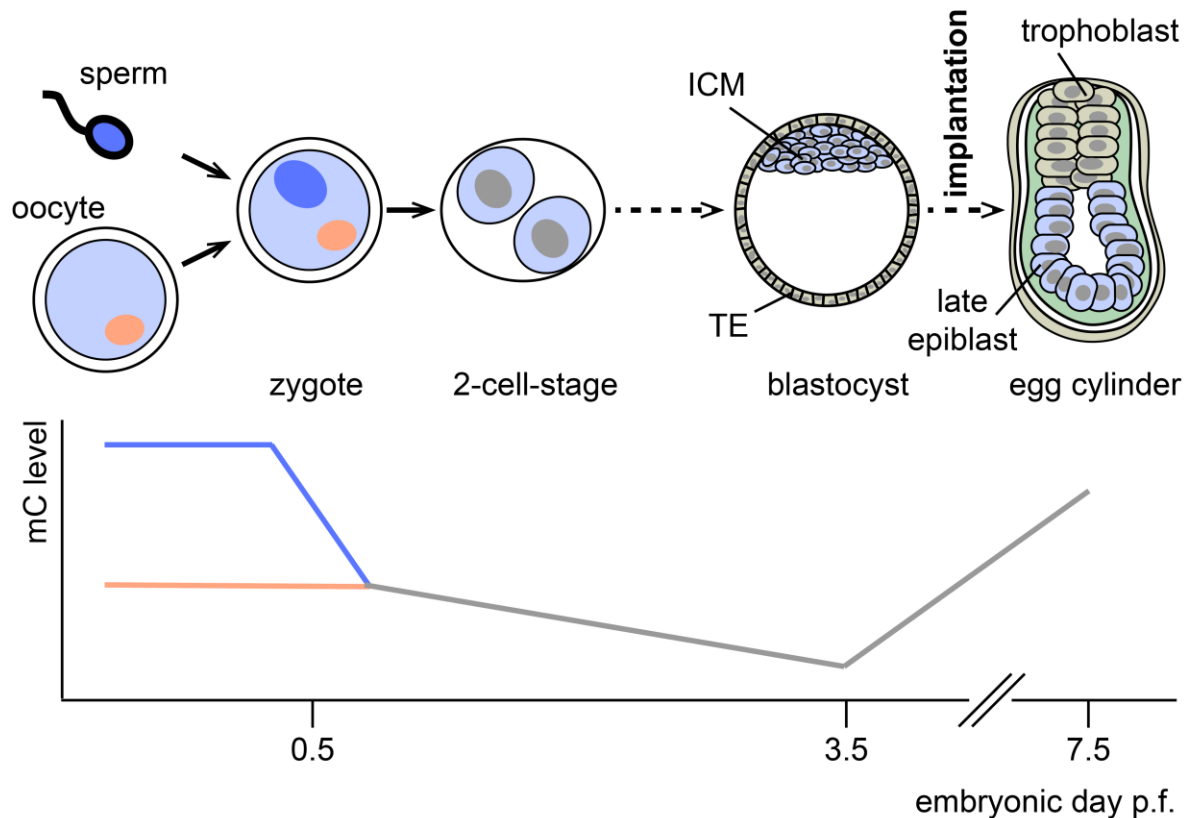
#### 1.3.1 Early *in vivo* development of mouse embryos

Epigenetic processes play a pivotal role in the development of the embryo where they govern differentiation into various mature tissues. The totipotent zygote, which is generated by fertilization, i.e. fusion of the haploid oocyte with a sperm, has the ability to develop into all different cell types, both embryonic and extraembryonic. After several uniform cell division cycles, cell polarity is established at the morula stage, followed by the first cellular differentiation in the blastocyst, which consists of two cell types: the Inner Cell Mass (ICM) and the trophectoderm (TE). Cells of the ICM are pluripotent and will give rise to the embryo proper, whereas TE cells will form the trophoblast and contribute to the placenta. The blastocyst hatches from its surrounding coat, the *zona pellucida*, and implants into the uterus while developing into the epiblast. The epiblast is usually divided into two stages: pre- and post-implantation epiblast, also referred to as early and late epiblast (Reik *et al.*, 2001; Simmons and Cross, 2005; Wang and Dey, 2006; Palini *et al.*, 2011).

Development of the zygote to the post-implantation embryo is a process that is accompanied by several massive changes in the epigenome, especially in DNA methylation (Auclair and Weber, 2012). A wave of global DNA demethylation occurs after fertilization and genomic DNA is remethylated during implantation, when cells from the ICM of the blastocyst develop into the epiblast (Smith *et al.*, 2012) (Figure 6). In the zygote, DNA demethylation occurs in an asymmetric manner: the paternal pronucleus rapidly loses cytosine methylation before the first cell division whereas the maternal genome gets gradually and passively demethylated over several rounds of DNA replication (Auclair and Weber, 2012). The loss of paternal cytosine methylation is accompanied by accumulation of hmC, generated specifically by TET3 (Gu *et al.*, 2011; Iqbal *et al.*, 2011). In contrast, the maternal pronucleus is protected from oxidation by binding of DPPA3 (Developmental pluripotency-associated protein 3, also known as Stella or PGC7) to H3K9me2, inhibiting TET3 recruitment (Wossidlo *et al.*, 2011; Nakamura *et al.*, 2012). However, TET3-mediated oxidation appears to be dispensable for embryonic development and demethylation of the paternal genome (Inoue *et al.*, 2015; Amouroux *et al.*, 2016) and the importance of this process is still debated.

Global DNA methylation levels of the embryo reach a minimum at the blastocyst stage (Smith *et al.*, 2012), at which mC is much more abundant in the ICM than in the TE (Dean *et al.*, 2001). During epiblast differentiation, the genome is remethylated within a few days (Borgel *et al.*,

2010). Global DNA methylation levels at the late epiblast already closely resemble those of fully differentiated tissues (Smith *et al.*, 2012). Establishment of these methylation patterns largely relies on the activity of the *de novo* methyltransferases DNMT3A and DNMT3B (Okano *et al.*, 1999; Auclair *et al.*, 2014).



**Figure 6: Early embryonic development**

Schematic representation of early embryonic development with an illustration of changes in genomic DNA methylation levels (orange: maternal genome, blue: paternal genome, gray: diploid embryonic genome). For the blastocyst and egg cylinder stage, mC levels for ICM and epiblast are depicted, respectively. ICM: inner cell mass, TE: trophectoderm, p.f.: post fertilization (Wang and Dey, 2006; Hemberger *et al.*, 2009; Auclair and Weber, 2012; Smith *et al.*, 2012).

### 1.3.2 *In vitro* models of differentiation

Murine ESCs are widely used to study pluripotency and differentiation as cell culture models *in vitro* (Nichols and Smith, 2012). ESC lines are derived from the ICM and have initially been cultured on so-called feeder cells, i.e. inactivated fibroblasts (Evans and Kaufman, 1981). When re-injected into the blastocyst and transplanted into the animal, ESCs can contribute to all tissues of a viable embryo including the germ line, demonstrating their pluripotency (Bradley *et al.*, 1984; Nichols and Smith, 2012).

Since cultivation of ESCs on feeder cells is labor-intensive and time-consuming, other culture systems have been developed. Leukemia inhibitory factor (LIF) was discovered a few years after

## 1. Introduction

the first establishment of ESC lines to inhibit cellular differentiation in the absence of feeder cells (Smith *et al.*, 1988; Williams *et al.*, 1988; Nichols and Smith, 2012). LIF directly promotes expression of core pluripotency factors such as OCT4, SOX2, NANOG, and others via activation of STAT3 (Nichols and Smith, 2012). A few years ago, an alternative medium was described that allows cultivation of ESCs not only in feeder-, but also in serum-free conditions. The combination of two small molecule inhibitors, generally referred to as "2i", blocks GSK3 (glycogen synthase kinase-3) and MAPK/ERK (mitogen-activated protein kinase) signaling and leads to ground state pluripotency of ESCs (Ying *et al.*, 2008). Interestingly, addition of 2i to the ESC culture causes DNA hypomethylation by downregulation of DNMT3A, DNMT3B, and DNMT3L. When transferring ESCs from serum/LIF conditions to 2i medium, this demethylation is accompanied by a transient peak in TET1 and TET2 expression and genomic hmC levels (Ficz *et al.*, 2013). Culturing ESCs under 2i conditions leads to a so-called "naïve" pluripotent state which resembles the ICM of the blastocyst or the early epiblast, whereas serum/LIF culture induces a more "primed" pluripotent state that shows similar properties to the late epiblast (Habibi *et al.*, 2013; Leitch *et al.*, 2013). There are several culture systems that allow differentiation of ESCs to distinct cell types. For example, withdrawal of 2i/LIF and addition of Activin A and FGF2 to the medium leads to formation of epiblast-like cells (EpiLCs) resembling post-implantation embryonic cells. Similar to the *in vivo* situation, this process is accompanied by downregulation of pluripotency genes such as NANOG, OCT4, and KLF4 and upregulation of the *de novo* DNMTs (Hayashi *et al.*, 2011). Another example is formation of a homogeneous neuronal cell population, which requires a more complex protocol, starting with ESCs cultured in serum/LIF conditions. An intermediate step of cellular aggregation is followed by addition of retinoic acid and culture on poly-DL-ornithine/laminin-coated surfaces, leading to generation of neuronal progenitor cells (NPCs). NPCs are characterized by expression of NESTIN and can further develop into mature neurons (Bibel *et al.*, 2004; Bibel *et al.*, 2007). Since massive changes of chromatin structure and gene transcription occur during such cellular differentiation processes, *in vitro* models of differentiation represent an invaluable tool to study epigenetic regulation and dynamics.

## 2. Results

### 2.1 Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives

published February 28, 2013, in *Cell*

# Dynamic Readers for 5-(Hydroxy)methylcytosine and Its Oxidized Derivatives

Cornelia G. Spruijt,<sup>1,9</sup> Felix Gnerlich,<sup>2,9</sup> Arne H. Smits,<sup>1</sup> Toni Pfaffeneder,<sup>2</sup> Pascal W.T.C. Jansen,<sup>1</sup> Christina Bauer,<sup>3</sup> Martin Münzel,<sup>2</sup> Mirko Wagner,<sup>2</sup> Markus Müller,<sup>2</sup> Fariha Khan,<sup>4,5</sup> H. Christian Eberl,<sup>6</sup> Anneloes Mensinga,<sup>1</sup> Arie B. Brinkman,<sup>7</sup> Konstantin Lephikov,<sup>8</sup> Udo Müller,<sup>3</sup> Jörn Walter,<sup>8</sup> Rolf Boelens,<sup>5</sup> Hugo van Ingen,<sup>5</sup> Heinrich Leonhardt,<sup>3</sup> Thomas Carell,<sup>2,\*</sup> and Michiel Vermeulen<sup>1,\*</sup>

<sup>1</sup>Department of Molecular Cancer Research, Proteomics and Chromatin Biology, UMC Utrecht, 3584 CG Utrecht, the Netherlands

<sup>2</sup>Center for Integrated Protein Science at the Fakultät für Chemie und Pharmazie, Ludwig-Maximilians-Universität München, 81377 Munich, Germany

<sup>3</sup>Center for Integrated Protein Science at the Fakultät für Biologie, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany

<sup>4</sup>University Institute of Biochemistry and Biotechnology, Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, Rawalpindi, Pakistan

<sup>5</sup>NMR Spectroscopy Research Group, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands

<sup>6</sup>Proteomics and Signal Transduction, Max-Planck-Institut für Biochemie, 82152 Martinsried, Germany

<sup>7</sup>Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen, 6525 GA Nijmegen, the Netherlands

<sup>8</sup>Genetik/Epigenetik, Universität des Saarlandes, 66123 Saarbrücken, Germany

<sup>9</sup>These authors contributed equally to this work

\*Correspondence: thomas.carell@lmu.de (T.C.), m.vermeulen-3@umcutrecht.nl (M.V.)

<http://dx.doi.org/10.1016/j.cell.2013.02.004>

## SUMMARY

Tet proteins oxidize 5-methylcytosine (mC) to generate 5-hydroxymethyl (hmC), 5-formyl (fC), and 5-carboxylcytosine (caC). The exact function of these oxidative cytosine bases remains elusive. We applied quantitative mass-spectrometry-based proteomics to identify readers for mC and hmC in mouse embryonic stem cells (mESC), neuronal progenitor cells (NPC), and adult mouse brain tissue. Readers for these modifications are only partially overlapping, and some readers, such as Rfx proteins, display strong specificity. Interactions are dynamic during differentiation, as for example evidenced by the mESC-specific binding of Klf4 to mC and the NPC-specific binding of Uhrf2 to hmC, suggesting specific biological roles for mC and hmC. Oxidized derivatives of mC recruit distinct transcription regulators as well as a large number of DNA repair proteins in mouse ES cells, implicating the DNA damage response as a major player in active DNA demethylation.

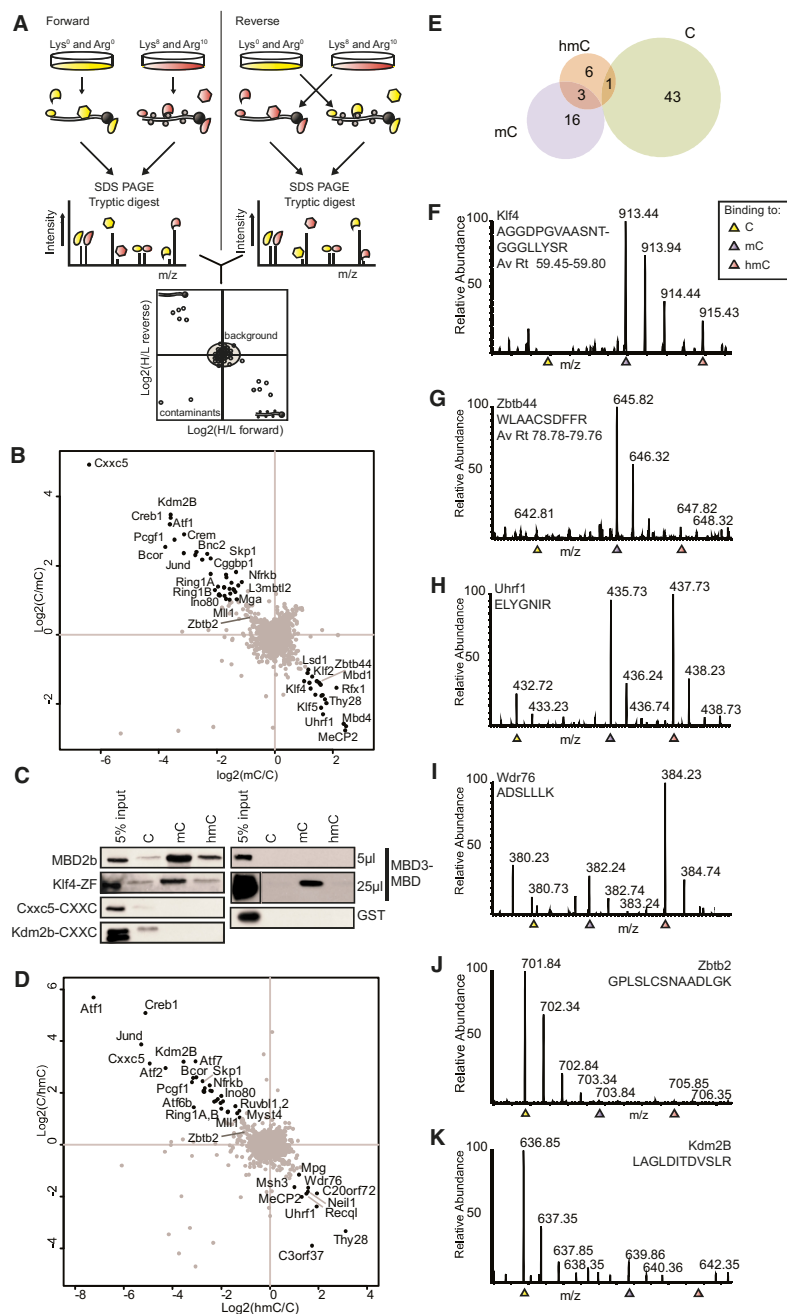
## INTRODUCTION

Methylation of cytosine residues at carbon atom 5 of the base (mC) represents a major mechanism via which cells can silence genes. Cytosine methylation mostly occurs in a CpG dinucleo-

tide context. However, CpG islands (CGIs), which are characterized by a very high CpG density and are often found in promoter regions of genes, are typically hypomethylated. Methylation of these CGIs results in transcriptional silencing. The molecular mechanisms underlying the association between DNA methylation and repression of transcription have proven difficult to decipher. The classic view is that methylation of DNA results in the recruitment of methyl-CpG-binding proteins (MBPs) that possess transcriptionally repressive enzymatic activities (Defossez and Stancheva, 2011). However, in vivo validation for this model on a genome-wide level is still lacking. In contrast, recent in vivo data have revealed that CXXC-domain-containing proteins specifically bind to nonmethylated cytosines. In this case, hypomethylated CGIs serve as a recruitment signal for CXXC-domain-containing activators that establish a transcriptionally active chromatin state (Thomson et al., 2010).

It was discovered 4 years ago that Tet enzymes convert mC to 5-hydroxymethylcytosine (hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). This modification is particularly abundant in the brain and in embryonic stem cells but is detectable in all tissues tested (Globisch et al., 2010; Szwagierczak et al., 2010). Tet enzymes can catalyze further oxidation of hmC to 5-formylcytosine (fC) and 5-carboxylcytosine (caC) (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011). fC and caC can subsequently serve as substrates for thymine-DNA glycosylase (Tdg), which eventually results in the generation of a nonmethylated cytosine (He et al., 2011; Maiti and Drohat, 2011). Therefore, this Tet-Tdg pathway represents an active DNA demethylation pathway. It is not clear whether hmC, fC, and caC have additional DNA-demethylation-independent





**Figure 1. Identification of mC- and hmC-Specific Readers in Mouse Embryonic Stem Cells**

(A) Schematic overview of the workflow. (B) Scatterplot of a SILAC-based mC DNA pull-down in mESC nuclear extracts.

(C) Validation of the mC-specific binding of Klf4 and nonmethyl-C-specific binding of Cxxc5 and Kdm2b. DNA pull-downs were performed with recombinant GST-fusion proteins followed by western blotting. For MBD3\_25, an empty lane was cut out.

(D) Scatterplot of a SILAC-based hmC DNA pull-down in mESC nuclear extract.

(E) Venn diagram showing overlap of readers for C, mC, and hmC.

(F–L) Representative mass spectra obtained in the triple-SILAC DNA pull-down in mESCs. Each spectrum shows the relative affinity of the indicated peptides and proteins for nonmethylated (yellow), methylated (blue), and hydroxymethylated (red) DNA.

See also Figure S1 and Table S1.

data reveal that each cytosine modification recruits a distinct and dynamic set of proteins. The known biology of these interacting proteins suggests a role for hmC, fC, and caC in active DNA demethylation pathways via base excision repair (BER), as well as an epigenetic recruitment function in certain cell types.

## RESULTS

### Identification of mC and hmC Readers in mESCs

To identify readers for methylcytosine and its oxidized derivatives, we made use of a DNA pull-down approach combined with quantitative MS. In brief, nuclear extracts from mESCs grown in “light” or “heavy” SILAC medium were incubated with a nonmodified or modified double-stranded DNA sequence (5′-AAG.ATG.ATG.AXG.AXG.AXG.AXG.ATG.ATG-3′, with X representing C, mC, or hmC; “forward” pull-down; Figure 1A). As a control, a label-swap, or “reverse,” experiment was performed. Following incubation and washes, beads were combined and bound proteins were in-gel digested

with trypsin and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Raw MS data were analyzed using MaxQuant (Cox and Mann, 2008). Specific interactors are distinguishable from background proteins by their H/L ratio. Proteins binding selectively to the modified DNA have a high ratio in the forward pull-down and a low ratio in the reverse pull-down, whereas readers for the nonmodified DNA show opposite

functions, as very few specific binders, or “readers,” for these oxidized versions of mC have been described thus far. We applied quantitative mass spectrometry (MS)-based proteomics to identify a large number of readers for mC and its oxidized derivatives in mouse embryonic stem cells (mESCs). Furthermore, we also identified readers for mC and hmC in neuronal progenitor cells (NPCs) and in adult mouse brain. Our

with trypsin and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Raw MS data were analyzed using MaxQuant (Cox and Mann, 2008). Specific interactors are distinguishable from background proteins by their H/L ratio. Proteins binding selectively to the modified DNA have a high ratio in the forward pull-down and a low ratio in the reverse pull-down, whereas readers for the nonmodified DNA show opposite

binding (low forward ratio, high reverse ratio). Background proteins will have a ~1:1 ratio in both pull-downs (Figure 1A).

As shown in Figure 1B and Table S1 available online, we identified 19 proteins enriched for mC compared to C in mESC nuclear extracts ( $p < 0.05$  and ratio  $> 2$  in both pull-downs). Among these are the methyl-CpG-binding proteins MeCP2, Mbd1, Mbd4, and Uhrf1 (Defossez and Stancheva, 2011). Other interactors include Rfx1 and Zfx3, which were previously identified as mC readers (Bartke et al., 2010; Sengupta et al., 1999). Interestingly, three Klf proteins were identified as mC readers: Klf2, -4, and -5. These proteins carry three Krüppel-like zinc fingers, just like the Kaiso family of mC-binding proteins. Klf4 is one of the four Yamanaka reprogramming factors and has not been previously identified as a mC-binding protein in HeLa or U937 cells (Bartels et al., 2011; Bartke et al., 2010). This may be due to the low expression of Klf4 in differentiated cells relative to mESCs. We confirmed the direct binding of the Klf4 Krüppel-like zinc fingers to mC using recombinant protein and two different DNA sequences (Figure 1C and S1A). A motif bearing similarities to a recently published consensus binding site for Klf4, as determined by ChIP-seq (GGGXGTG) (Chen et al., 2008), revealed that Klf4 binds this motif with the highest affinity when “X” is mC (Figure S1A). These results establish Klf4 as a sequence-specific mC binding protein.

Mining published bisulfite sequencing data of mESCs and NPCs (Stadler et al., 2012) and overlapping this data with the Klf4 ChIP-seq profile in mESCs (Chen et al., 2008) revealed a substantial number of methylated Klf4-binding sites in this cell type (Figure S1B), which are mainly intronic and intergenic (Figure S1C). Out of the 7,321 Klf4-binding sites in mESCs that were covered in the bisulfite sequencing data set, 1,356 show high levels of DNA methylation in mESCs (18.5%). Many of these Klf4-binding sites contain a methylated Klf4-binding motif, such as GGCGTG (Figures S1D and S1E). Interestingly, many Klf4-binding sites that are nonmethylated in ES cells become hypermethylated in NPC cells (Stadler et al., 2012) (Figures S1B and S1D). This finding may be highly relevant in the context of Klf4-mediated cellular reprogramming. During reprogramming, Klf4 may be able to bind these methylated loci in differentiated cells to initiate stem-cell-specific gene expression patterns. Enrichment analyses for functional domains among the mC interactors revealed DNA-binding zinc fingers to be significantly enriched (Benj.Hoch.FDR =  $10^{-2.45}$ ; Figure S3A). These zinc fingers may also interact with the methylated DNA in a sequence-specific manner.

In addition to the cluster of mC-binding proteins, a large number of proteins displayed preferential binding to nonmethylated DNA (Figure 1B, upper-left quadrant). Consistent with previous observations, this cluster of proteins contains a number of CXXC-domain-containing proteins that are known to preferentially bind to nonmethylated CpGs (Blackledge et al., 2010; Thomson et al., 2010). Examples include Cxxc5, Kdm2b, and Mll1 (see Figure 1C). We also identified other subunits of the Mll1 and PRC1.1 (Bcor) complexes, which most likely bind to the nonmethylated DNA indirectly via Mll1 and Kdm2b, respectively. Other interactors include the Ino80 chromatin-remodeling complex and zinc-finger-containing transcription factors such as Zbtb2, as well as basic leucine zipper-containing proteins

(enriched Benj.Hoch.FDR =  $10^{-5.57}$ ; Figure S3A) such as JunD, Creb1, and Atf7, for which sequence-specific DNA binding is most likely abolished by DNA methylation.

Readers for hmC showed partial overlap with proteins observed to interact with mC (Figure 1D, lower-right quadrant, and Figure 1E), as only three proteins interacted with both modified baits: MeCP2, Uhrf1, and Thy28. Uhrf1 and MeCP2 are known to bind both mC and hmC, although MeCP2 clearly binds with a higher affinity to mC compared to hmC (Frauer et al., 2011; Hashimoto et al., 2012; Mellén et al., 2012). Thy28 is an uncharacterized protein that is associated with apoptosis (Toyota et al., 2012) and contains an EVE domain, which is possibly involved in (ds)RNA binding (Bertonati et al., 2009). Interestingly, two DNA glycosylases (Mpg and Neil3) and a helicase (Recql) were identified as hmC readers in mESCs. These proteins might be involved in active DNA demethylation pathways to convert hmC back to cytosine via base excision repair mechanisms, as has been suggested previously (Hajkova et al., 2010; Wossidlo et al., 2010). In addition, a number of previously uncharacterized proteins, Wdr76 and C3orf37, preferentially bound to hmC compared to C. We purified WDR76 as a GFP fusion protein from HeLa cells and found interactions with OCR, HELLS, and GAN (Figure S1F). The mouse protein Hells, or Lsh, is a DNA helicase that has previously been implicated in regulating DNA methylation levels in cells (Dennis et al., 2001). Interestingly, OCR, or Spindlin-1, is a protein that is known to bind trimethylated H3 lysine 4 (H3K4me3) (Bartke et al., 2010). A large number of proteins preferentially bound to the nonmodified DNA, as was observed for the mC pull-down (Figure S1G). We validated some of these findings using western blotting for endogenous proteins (Figure S1H).

To further investigate the relative affinity of proteins for C versus mC versus hmC in a single experiment, we made use of a triple pull-down approach (Vermeulen et al., 2010), in which mESCs are grown in three different SILAC media. “Light,” “medium,” and “heavy” nuclear extracts derived from these cells are incubated with C-, mC-, and hmC-containing DNA, respectively (Table S1). Quantitative MS is used to visualize the relative abundance of a protein in each of the three different pull-downs. This experiment confirmed most of the observations made in Figures 1B and 1D, although for some proteins, the ratios in the triple pull-down are lower. As shown in Figures 1F and 1G, Klf4 and Zbtb44 preferentially bind to the methylated DNA. Other proteins bind to both modified baits, such as Uhrf1 (Figure 1H). Kdm2b preferentially binds to the nonmodified DNA (Figure 1K). Contrary to a previous report (Yildirim et al., 2011), we did not observe a specific interaction between MBD3 and hmC (forward ratio, 0.448; reverse ratio, 1.823). We validated these observations using recombinant protein (Figure 1C). At higher concentrations of recombinant MBD3 protein, we observed a specific interaction with mC (Figure 1C), which is in agreement with a recent study that revealed that MBD3 has the highest affinity for mC compared to hmC and C (Hashimoto et al., 2012).

Taken together, these experiments reveal that mC and hmC both recruit distinct proteins in mESCs with little overlap. Furthermore, a large number of proteins preferentially bind to nonmodified DNA. The number of observed interactions with

hmC is moderate, and some of these suggest that hmC acts as an intermediate in active DNA demethylation pathways in mESCs.

### fC and caC Recruit a Large Number of Proteins in Mouse Embryonic Stem Cells, Including DNA Glycosylases and Transcription Regulators

We also applied our SILAC-based DNA pull-down approach to identify readers for fC and caC in mESCs. Colloidal blue analysis revealed that the total amount of protein binding to each bait is similar (Figure S2A). Ratios of the forward and reverse pull-downs with hmC, fC, or caC were individually averaged, and these average ratios were then plotted against each other in two-dimensional graphs (Figures 2A–2C and Table S1). From these plots, it is clear that both fC (blue, purple, and green) and caC (yellow and green) recruit many more proteins than hmC does (red and purple). Strikingly, there is only limited overlap between fC and caC binders (green) (Figure 2D). One of the proteins that binds to fC and caC, but not to hmC, is Tdg, which is consistent with its reported substrate specificity (Maiti and Drohat, 2011). We validated this binding behavior using recombinant protein in electrophoretic mobility shift assays (EMSA) (Figures 2E and 2F). We also purified GFP-Tdg from ES cells to identify Tdg interaction partners (Figure S2B and Table S1). None of the Tdg interactors were identified as specific readers in the fC and caC pull-down, indicating that these fC and caC interactions are Tdg independent. Another fC-specific reader is the p53 protein, which plays an important role in DNA damage response (Kastan et al., 1991). Interestingly, Dnmt1 specifically interacted with caC. This interaction was confirmed by EMSA as well as western blotting using an antibody against endogenous protein (Figures 2F and S2C). We also identified subunits of the Swi/Snf chromatin-remodeling complex, such as BAF170, as readers for caC. Three proteins bind to all oxidized derivatives of mC: Thy28, C3orf37, and Neil1. GO term enrichment for biological processes shows that fC significantly enriches for proteins that are related to DNA repair (Benj.Hoch.FDR =  $10^{-2.71}$ ) (Figure S3A), whereas caC interactors are not enriched for any biological process. RNA-binding proteins, mitochondrial proteins, and other proteins that are less likely to be associated with regulation of gene expression or DNA repair binding were identified as binders for fC and caC (Table S3). Some of these may have a basic affinity for the formyl and carboxyl groups on the DNA strands, which are more reactive than methyl or hydroxymethyl. To exclude the possibility that many fC and caC interactors are binding to damaged or abasic DNA, we validated the homogeneity of the DNA strands using HPLC (Figure S2D). Furthermore, we analyzed the DNA before (blue) and after incubation (red) with mESC nuclear extract by MALDI-TOF-MS (Figure S2E). Quantification of the modified residues by LC-MS/MS shows that there is no significant loss of the modified bases after incubation with nuclear extract (Figure S2F). Figures 2A–2C also show that the group of proteins that bind preferentially to nonmodified cytosine (black, lower-left quadrant) shows a large overlap between the three pull-downs and contains the PRC1.1, Mll1, and Ino80 complexes. To compare the relative affinity of proteins for these three modifications in a single experiment, we performed a triple pull-down. Analyses of the triple pull-down ratios

for the identified fC and caC readers show similar trends, although some of the observed ratios are less prominent. As shown in Figures 2G–2L (and Table S1), the representative spectra of the indicated peptides of Tdg, Neil3, Mpg, Dnmt1, MeCP2, and Uhrf1 show relative ratios that are in agreement with ratios obtained in the independent experiments shown in Figures 2A–2C.

In summary, our data suggest that oxidized cytosine bases may induce a DNA damage response and trigger base excision repair pathways, which may finally result in DNA demethylation. In addition, each of these modifications recruits transcription regulators and other proteins that are not likely to be related to active DNA demethylation.

### NPCs Contain a Distinct Set of mC and hmC Readers, Including Uhrf2, which Has the Highest Affinity for hmC

To investigate whether interactions with mC and hmC are dynamic during differentiation, we differentiated mESCs to NPCs. Nuclear extracts were generated from these cells followed by DNA pull-downs. Because no SILAC-compatible neurobasal medium is available, these experiments were performed using label-free quantification (LFQ) (Eberl et al., 2013; Hubner and Mann, 2011). Each DNA pull-down is analyzed separately and in triplicate. For all of the identified proteins (Table S1), we used ANOVA statistics ( $p = 0.025$  and  $S_0 = 2$ ) to compare the relative enrichment of proteins for each of the three baits. All significant outliers (192) were hierarchically clustered based on correlation after normalization by row mean subtraction (Figure 3A). Protein enrichment is indicated in red, whereas lack of enrichment is shown in blue. A large number of proteins bind to C or mC, whereas fewer proteins are specifically enriched in the pull-downs with hmC. Three smaller groups of proteins bind specifically to two of the baits (C/hmC, C/mC, or mC/hmC). As was observed in the DNA pull-downs from mESC nuclear extracts, CXXC-domain-containing proteins (Kdm2b and Mll, indicated in black) and their associated factors Bcor/Ring1a/b (blue) and Rbbp5/Ash2l (black) are enriched in the DNA pull-downs with nonmodified DNA relative to mC- and hmC-containing DNA. We identified Mbd2 and associated Mi-2/NuRD complex subunits as mC readers (indicated in yellow). Other identified MBD proteins include Mbd4, MeCP2, and Mbd1. Furthermore, a number of winged-helix (WH)-domain-containing proteins bound specifically to mC, including Rfx5 and its associated factors Rfxap and Rfxank (orange), which have previously been identified as methyl CpG interactors (Bartke et al., 2010).

Strikingly, these proteins bind more strongly to C compared to hmC. We further substantiated these observations by using recombinant protein (Figure 3B). This result indicates that, for some readers, oxidation of mC not only weakens the interaction, but also repels the mC interactor. The homeobox domain is significantly enriched in the cluster of mC-specific readers (Benj.Hoch.FDR =  $10^{-1.8}$ , Figure S3A), which is consistent with a previous study (Bartke et al., 2010). In addition, several known mC readers, such as Kaiso, Uhrf1, and Mbd4, bind both modified forms of cytosine. A number of DNA glycosylases bind specifically to hmC (Neil1, Neil3), as well as some helicases (Hells, Harp, Recql, and its homolog Bloom), which again suggests a



(F) Electrophoretic mobility shift assays as shown in (E) performed with GFP-Tdg and GFP-Dnmt1 for all six residue variants (C, mC, hmC, fC, caC, and abasic site [AB]) in direct comparison to unmodified DNA. The binding preference was determined as the ratio of fluorescence signals of the different DNA substrates in the shifted bands. Shown are the means of three experiments; error bars represent SD.



DNA-repair-involved DNA demethylation pathway (GO DNA repair; Benj.Hoch.FDR =  $10^{-3.91}$ ; Figure S3A). Although homeobox proteins are known to bind specifically to mC, a number of homeobox proteins show preferential binding to hmC in NPC extracts (examples include Zhx1 and -2). Finally, Uhrf2 was identified as a specific hmC-binding protein in NPCs, which we confirmed using recombinant protein (Figure 3B). Uhrf2 is not expressed in mESCs, and its levels increase upon differentiation (Pichler et al., 2011). This explains why Uhrf2 was not identified as an hmC-specific reader in mESC DNA pull-downs.

Taken together, these experiments reveal that interactions with mC and hmC are highly dynamic during differentiation. Furthermore, the observations made in NPCs strengthen our hypothesis that oxidation of mC serves as a trigger for active DNA demethylation. Nevertheless, some hmC-specific readers in NPCs do not appear to be linked to DNA repair mechanisms, indicating that, in these cells, hmC may also serve a role as a “classical” epigenetic mark that recruits transcriptional regulators.

#### NMR-Based Analysis of the Rfx5 WH Domain Bound to mC DNA

The specific interaction between the Rfx5 WH domain and mC DNA was studied in detail using solution nuclear magnetic resonance (NMR) spectroscopy in order to derive binding affinity and identify the mC-binding site. Addition of a singly methylated 18bp DNA fragment to the Rfx5-WH domain results in large changes in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC “fingerprint” spectrum (Figure 3C). After addition of a slight molar excess of DNA, the spectrum does not show any further changes, indicating that Rfx5-WH strongly binds mC DNA and preferentially at only one of the two mC sites (Figure 3C). The affinity of Rfx5 for mC DNA was derived from the observed peak displacement for residues in the fast exchange regime, such as T104 and E102, assuming that the two mC are independent and equivalent, which resulted in an apparent dissociation constant  $K_{D,app}$  of  $\sim 3 \mu\text{M}$  (with 95% probability limits  $10 \text{ nM} < K_D < 16 \mu\text{M}$ ) (Figure 3D and Supplemental Information). Based on DNA pull-downs done with recombinant protein, which revealed a quantitative depletion of the WH domain from the lysate, we anticipate the  $K_D$  to be in the nM range (Figure 3B). To identify the residues that are responsible for specific mC binding, we used the DNA-bound Rfx1 WH domain crystal structure (PDB ID 1DP7; sequence identity 35%; Avvakumov et al., 2008; Gajiwala et al., 2000) to construct a homology model structure of Rfx5-WH and validated it against the experimental chemical shifts (data not shown). The homology model contains a hydrophobic pocket that includes residues with the largest chemical shift changes and is well aligned with an extended basic surface that is responsible for DNA binding in Rfx1. This binding pocket, formed by the side chains of K110, V113, Y114, T132, F135, L139, and Y169, is appropriately shaped to capture the mC base via a flip-out

mechanism, as seen in the case of UHRF1 (Figure 3E). Steric clashes introduced by the presence of an additional hydroxyl group could cause the observed specificity for mC. Given the apparent high affinity and DNA-sequence-independent binding to mC, we propose that the WH domain that is present in Rfx proteins is a bona fide mCpG-binding domain.

#### Brain-Specific Readers for mC and hmC Include Dlx Proteins

The adult brain is the organ with the highest levels of hmC (Globisch et al., 2010). Tet enzymes and hmC have been shown to play a role in active DNA demethylation of certain genes in this organ (Guo et al., 2011). To identify readers for C, mC, and hmC in the adult brain, nuclear extracts were prepared from this tissue, and these extracts were used for DNA pull-downs. LFQ was used to determine differential binders (Table S1). In brain extracts, we identified fewer specific readers compared to NPCs (108,  $p = 0.025$  and  $S_0 = 0$ ; Figure 4), most likely due to the presence of highly abundant structural proteins derived from connective tissue and extracellular matrix in these nuclear extracts. Interestingly, more proteins specifically bind to hmC compared to mC in brain extracts. This is in contrast to NPCs and mESCs, in which more interactions with mC relative to hmC are observed, which may imply a specific role for hmC in brain tissue.

The nonmodified DNA pull-down enriched for the same factors as those observed in mESCs and NPCs, including Cxxc5, Kdm2b, and Bcor (CXXC-domains indicated in black, PRC1 complex in blue, and Ino80 in red). In this case, mC DNA was bound by the Mbd2/NuRD complex, which contains the brain-specific ATPase Chd5 (Eberl et al., 2013; Potts et al., 2011) (indicated in yellow). Interestingly, we identified three distal-less homeobox proteins (Dlx1, -5, and -6) as specific mC interactors. Dlx proteins play a role in the development of the brain and are also expressed in specific regions of the adult brain (Jones et al., 2011; Wang et al., 2011). Wdr76 and Thy28 are hmC specific, as was also observed in NPCs. Thap11 (or Ronin) is identified as a brain-specific hmC reader. Interestingly, this protein is highly expressed in certain regions of the brain, including all Purkinje cells (Dejosez et al., 2008). Finally, we identified all four subunits of replication factor C (Rfc2-5) and the associated factor Rfc1 as hmC-specific readers (indicated in green).

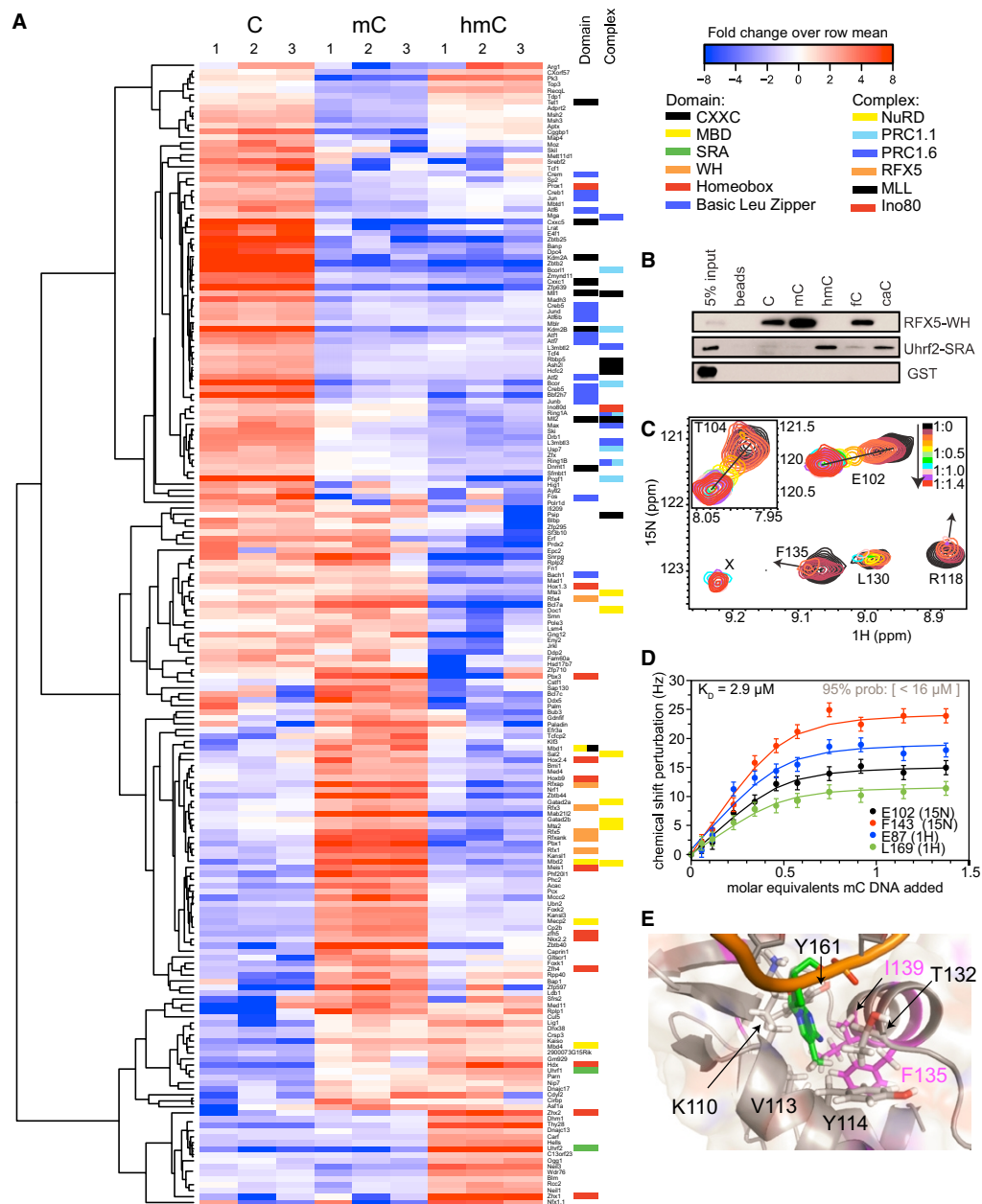
Altogether, these experiments further emphasize the dynamic nature of the mC and hmC interactomes during development.

#### Global Absolute Quantification of Protein Levels in mESCs, NPCs, and Adult Mouse Brain Extracts Reveals Expression-Level-Dependent and -Independent Interaction Dynamics

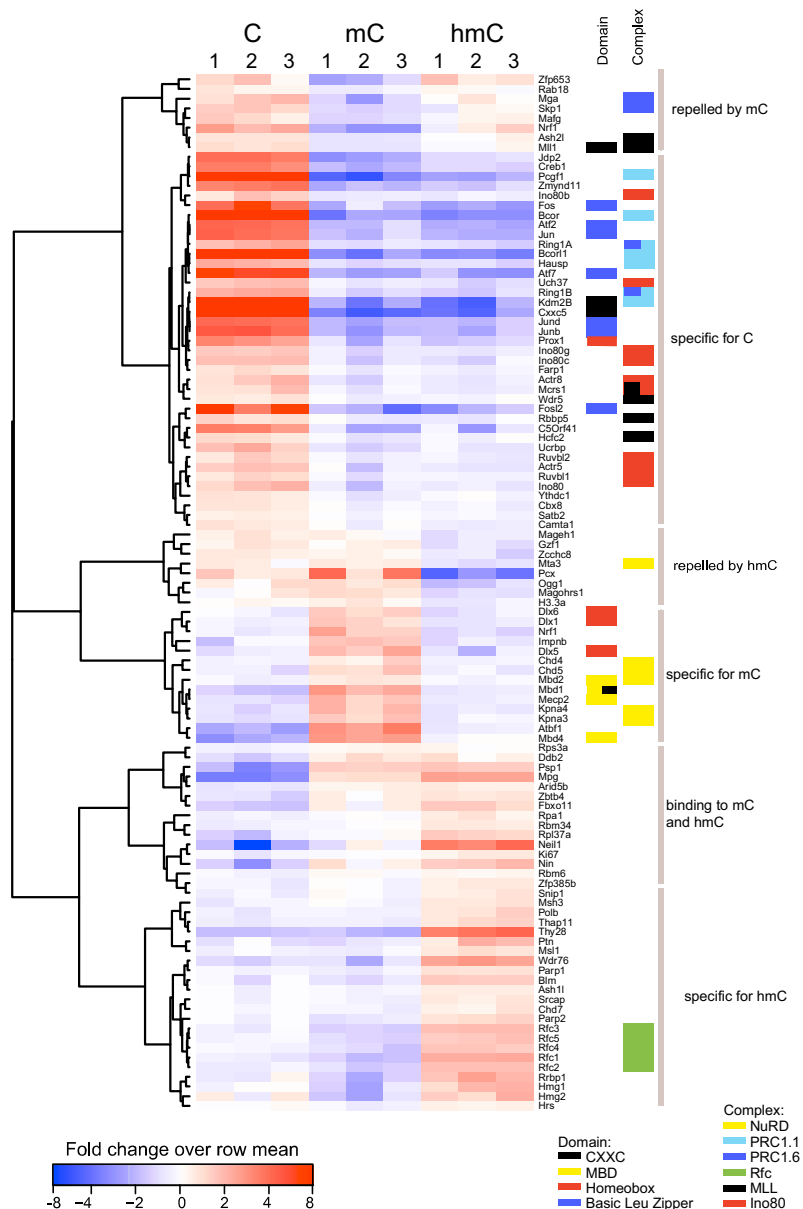
Our screening for mC- and hmC-specific readers in mESCs, NPCs, and adult mouse brain revealed a large number of cell-type- or organ-specific interactors (Figure S3B). The most

(G–L) Representative spectra of the indicated peptides obtained in the triple-labeled DNA pull-down in mESCs. Each spectrum shows the relative affinity of the indicated peptides and proteins for hmC-containing (red), fC-containing (blue), and caC-containing (yellow) DNA. Spectra are shown for Tdg (G), Neil3 (H), Mpg (I), Dnmt1 (J), MeCP2 (K), and Uhrf1 (L).

See also Figure S2 and Table S1.



**Figure 3. Hierarchical Clustering of NPC-Specific C, mC, and hmC Readers**  
(A) Correlation-based clustering of the LFQ intensities after log2 transformation and normalization by row mean subtraction. Included in the clustering are proteins that are significantly binding to at least one of the baits as determined by an ANOVA test. Blue indicates lack of enrichment, whereas enrichment is indicated in red. Domain and Complex columns indicate the DNA-binding domain(s) that may be responsible for direct binding to the bait and the complexes that readers are part of, respectively.  
(B) Biochemical validation experiments using DNA pull-downs with recombinant DNA-binding domains.  
(C) Overlay of Rfx5-WH HSQC spectra with increasing amounts of mC DNA added and color-coded on the indicated scale listing the WH domain:DNA ratio. Some residues, such as F135 and R118, cannot be unambiguously tracked to their bound states because their chemical shift changes are very large. Peaks corresponding to their bound state, such as "X," appear only after addition of a full molar equivalent of DNA.  
(D) Selected binding curves and fits for resonances that are in the fast exchange regime throughout the titration. Error bars (SD) for the peak positions are set to 1.2 Hz.  
(E) Close-up of the putative mC-binding pocket in the RFX5 WH domain. The methylated cytosine is indicated in green.  
See also Table S1.



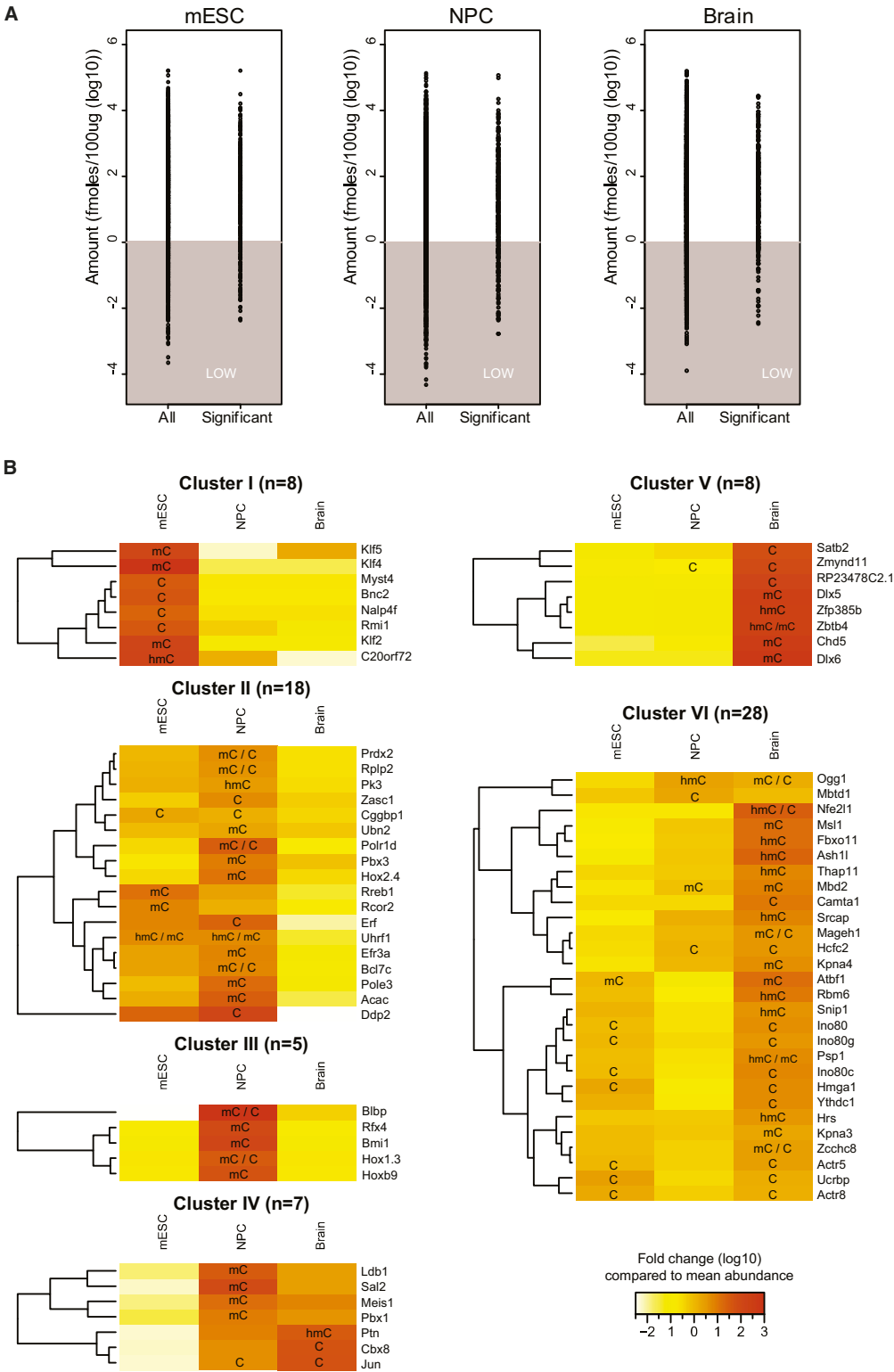
**Figure 4. Hierarchical Clustering of Brain-Specific C, mC, and hmC Readers**

Correlation-based clustering of the row-mean-subtracted LFQ intensities of proteins in C, mC, and hmC DNA pull-downs in adult mouse brain nuclear extracts. See also Table S1.

The cluster of mESC-specific proteins is enriched for anchoring junction (Benj.Hoch.FDR =  $10^{-2.96}$ ) and cell adhesion (Benj.Hoch.FDR =  $10^{-2.14}$ ), whereas proteins in brain enriched GO terms such as synaptic transmission (Benj.Hoch.FDR =  $10^{-3.77}$ ) and cognition (Benj.Hoch.FDR =  $10^{-2.75}$ ), as expected (Figure S4C). The molar concentrations of proteins that are significantly enriched in one of the DNA pull-downs are spread over several orders of magnitude, indicating that our screening is not biased toward high-abundant proteins (Figure 5A). Of the 259 proteins that showed dynamic interactions through development (Table S3), 20 proteins were not quantified in the iBAQ measurements. The 74 proteins (~31%) that do show a correlation between interaction pattern and protein abundance in the different extracts can be divided into six clusters (Figure 5B). A correlation was defined as gaining or losing an interaction accompanied by at least a 2-fold change in protein abundance. An example of a protein that was identified as a specific (mC) reader only in mESCs was Klf4. As shown in Figure 5B, this protein is highly expressed in mESCs but is less abundant in NPCs or in the adult mouse brain. Another example is represented by the Dlx5 and Dlx6 proteins, which are highly abundant in brain nuclear extract and exclusively bind to mC in pull-downs from these extracts. For about 185 proteins, no correlation is observed between expression levels (at least 2-fold change) and binding behavior. For these proteins, the

obvious explanation for these observed differential interactions is regulation of reader abundance at the protein level. Alternatively, the interaction between a reader and (modified) DNA may be affected by posttranslational modifications (PTMs). To investigate global absolute protein levels in the different nuclear extracts that were used for the pull-downs, we made use of a method called intensity-based absolute quantification (iBAQ) (Schwanhäusser et al., 2011). Approximately 8,000 proteins were quantified in at least one of the extracts (Table S2). All proteins with at least a 10-fold change in concentration were clustered based on their expression pattern (Figure S4B).

cause of differential binding may be explained through PTMs that affect the interaction between a reader and DNA or a differentially expressed cofactor. A good example of the latter is the Mi-2/NuRD complex. Although most of its subunits display equal expression levels in mESCs, NPCs, and brain, mC-specific interactions are not observed in mESCs. This can be explained by the fact that Mbd2, which is the direct reader of mC within the NuRD complex, is low in abundance in mESCs and is upregulated during differentiation (Figure 5B). Thereby, it controls the mC-specific binding of the entire complex. In mESCs, the majority of the Mi-2/NuRD complex contains Mbd3, which is



(legend on next page)



the MBD-containing protein that has lost its high-affinity mC binding ability. Furthermore, technical reasons for not identifying an interactor could be the presence of highly abundant structural proteins in the brain lysate or binding competition among different readers in the extracts. Altogether, the absolute quantification of protein abundance in the different nuclear extracts revealed large differences in protein levels between mESC, NPCs, and adult mouse brain. This data set serves as a rich resource on its own but also enables us to explain many of the differential interactions that we identified using quantitative MS-based interactomics.

### Uhrf2 Stimulates the Sequential Activity of the Tet1 Enzyme

The first protein that was identified as an hmC binder was Uhrf1 (Frauer et al., 2011), a protein that is involved in maintenance of DNA methylation (Bostick et al., 2007). Our data revealed that Uhrf1 binds with a similar affinity to mC and hmC, which is consistent with previously published data (Frauer et al., 2011). This is in contrast to Uhrf2, which we identified as a high-affinity hmC-binding protein in NPC cells that shows a lower affinity for mC. The function of Uhrf2 is not well understood. It is clear, however, that Uhrf2 cannot rescue the phenotype of Uhrf1 knockout cells, which lose DNA methylation (Pichler et al., 2011; Zhang et al., 2011). Uhrf1 is highly expressed in mESCs, whereas Uhrf2 levels increase during differentiation (Table S3 and Pichler et al., 2011). Altogether, this prompted us to investigate whether Uhrf2 expression affects the levels of mC and its oxidized derivatives. The Tet1-catalytic domain was transfected into HEK293T cells with and without coexpression of Uhrf2. Total genomic DNA modification levels were determined using LC-MS/MS (Figure 6 and Supplementary Information). As shown in Figure 6D, Uhrf2 overexpression increases the level of hmC. More striking is the increase of fC and caC levels upon Uhrf2 coexpression together with the Tet1 catalytic domain. Because fC and caC serve as substrates for Tdg and BER, the detected increase in the levels of fC and caC following Uhrf2 expression may be an underestimation of the actual production of these bases. It therefore seems that Uhrf2 promotes repetitive oxidation of mC by the Tet proteins. We hypothesize that flipping the modified cytosine base out of the DNA double helix, as has been described for Uhrf1 binding to methylated and hydroxymethylated DNA (Arita et al., 2008; Frauer et al., 2011), may enhance accessibility of the hydroxymethylated base to the Tet enzymes, thereby promoting further oxidation.

### DISCUSSION

In this study, we have used quantitative MS-based proteomics to identify readers for mC and its oxidized derivatives in mESCs,

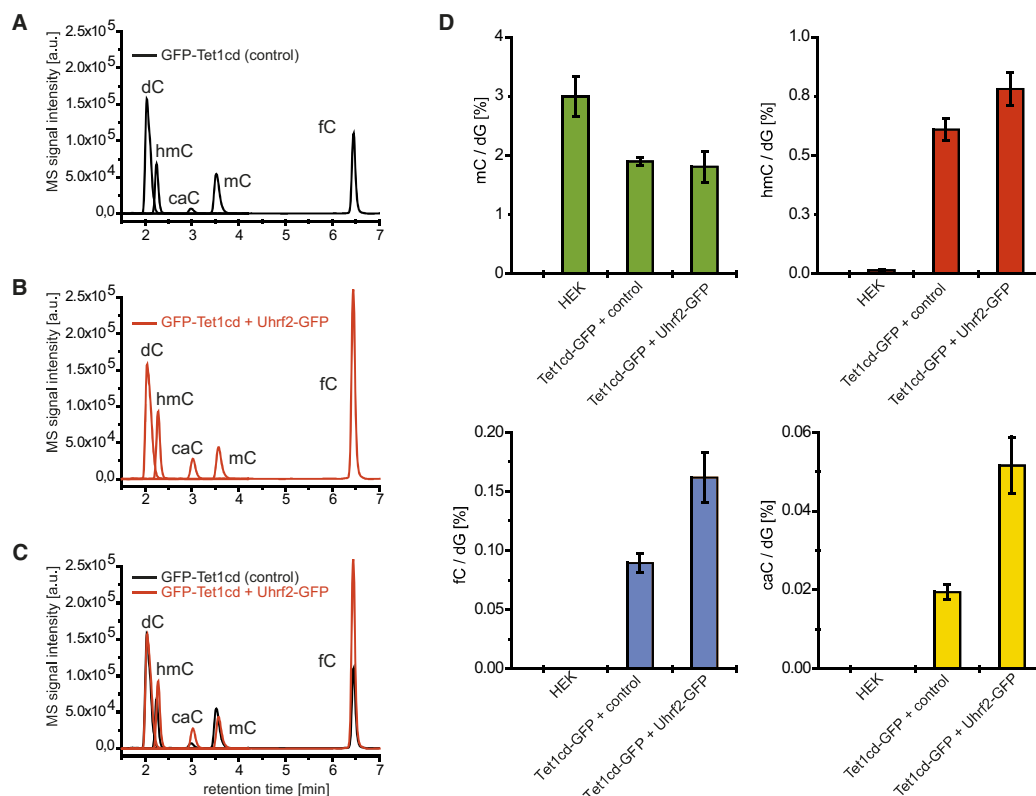
as well as readers for mC and hmC in NPCs and adult mouse brain. Readers for individual modifications were found to be highly dynamic throughout the three cell types and tissues that we investigated (Figure 7). This is in contrast to interactions with histone modifications, such as trimethylated lysines on histone H3. For these modifications, the majority of interactors are constant between different cell types or developmental stages (Eberl et al., 2013 and M.V., unpublished data). Readers for distinct cytosine modifications show limited overlap. This indicates that, at least from a biochemical perspective, mC, hmC, fC, and caC behave quite differently. Although little overlap was observed with regard to proteins that interact with each of the epigenetic marks, they all repelled a common set of proteins, such as several CXXC-domain-containing proteins and their interactors. It remains to be determined which of the consequences of DNA (hydroxy)methylation is functionally most relevant: recruitment of transcriptionally repressive complexes or preventing the binding of certain (activating) proteins to unmodified DNA. A detailed biochemical characterization of the interactions and their dissociation constants will be important to answer this question.

Our experiments revealed a number of DNA glycosylases and DNA repair proteins that bind to hmC, fC, and caC, whereas we identified few such proteins binding to mC. The enriched binding of DNA-repair-associated proteins was most pronounced for fC. From this observation, one can conclude that the conversion of hmC to fC is a signal that is likely to result in repair-associated removal of the modified base by proteins that are rather ubiquitously expressed. It is therefore surprising that, in different cell types and tissues, rather constant levels of hmC, fC, and caC are found. The maintenance of such constant levels of these bases in mESCs may indicate a high turnover of DNA methylation, probably involving a constant “correction” by *de novo* methylation. Regardless, it will be important to investigate which mechanisms control Tet enzyme conversion of mC to hmC and further oxidation to fC and caC. Our data reveal that coexpression of Uhrf2 with the catalytic domain of Tet1 results in a (transient) upregulation of hmC, fC, and caC, indicating that Uhrf2 promotes the sequential oxidation of mC by Tet1. One of the other factors influencing the catalytic activity of the Tet enzymes is the concentration of cellular metabolites. It has been shown that oncometabolites such as 2-hydroxyglutarate can competitively inhibit the activity of 2-oxo-glutarate-dependent enzymes, such as the Tet proteins (Chowdhury et al., 2011; Xu et al., 2011). Furthermore, mutations in IDH1 and -2, which generate 2-oxo-glutarate, are phenocopied by mutations in the TET enzymes and result in cancer (Figuerola et al., 2010). Mutations in the IDH2 and TET2 genes were also linked to lower genomic hmC levels and altered gene expression patterns in myeloid cancers (Ko et al.,

### Figure 5. Global Absolute Protein Quantification in mESCs, NPCs, and Adult Mouse Brain

(A) Graphs indicating the concentration of all proteins identified in the nuclear extract (all) and the identified readers (significant) in each of the cell types. The gray area indicates the concentration at which protein quantification is inaccurate. (B) Readers for which protein expression levels correlate with DNA binding patterns were clustered into six groups based on their expression in the three different nuclear extracts. The color indicates protein levels (white, low; red, high), whereas binding preference is indicated by C, mC, hmC, or combinations thereof.

See also Figure S4 and Table S2 and Table S3.



**Figure 6. Stable Isotope-Dilution-Based LC-ESI-MS/MS Quantification of Cytosine Derivatives in HEK293T Cells**

(A) Nonquantitative LC-MS/MS chromatogram of digested genomic DNA from HEK293T cells cotransfected with Tet1-catalytic domain-GFP (GFP-Tet1cd) and an unrelated expression construct (control). Depicted are the overlaid ion chromatograms of the MS/MS transitions for dC and the cytosine derivatives (black curves). dC, mC, and hmC were measured by a factor of  $\sim 10^2$ – $10^3$  less sensitive in comparison to caC and fC.

(B) Same as (A) except that Uhrf2-GFP was coexpressed together with GFP-Tet1cd. The MS signal intensities were normalized to the dC content of (A).

(C) Superposition of (A) and (B).

(D) Levels of cytosine derivatives relative to the total cytosine content (dG) as determined by quantitative LC-MS/MS mass spectrometry.

Shown are the means of technical triplicates; error bars reflect SD.

2010; Konstandin et al., 2011). In support of these observations, which clearly link hmC to cancer, we noticed that many hmC, fC, and caC readers are implicated in cancer, including UHRF2, CARF, p53, and HELLS (Lee et al., 2000). Interestingly, mutations in the Hells helicase, which we identified as an hmC reader in NPCs, result in a decrease of DNA methylation levels in cells (Myant et al., 2011). It seems clear that regulating the levels of mC and its oxidized derivatives is essential for normal cell homeostasis and that deregulation of the readers, writers, and erasers of these marks results in a disturbance of the balance between cell proliferation and differentiation during development.

## EXPERIMENTAL PROCEDURES

### Cell Culture

IB10 mESCs were cultured in light ( $R^0K^0$ ) or heavy ( $R^{10}K^8$ ) SILAC medium in the presence of 2i compounds. For triple labeling, a third type of medium was used containing medium-labeled L-lysine ( $K^4$ ) and L-arginine ( $R^6$ ). mESCs were

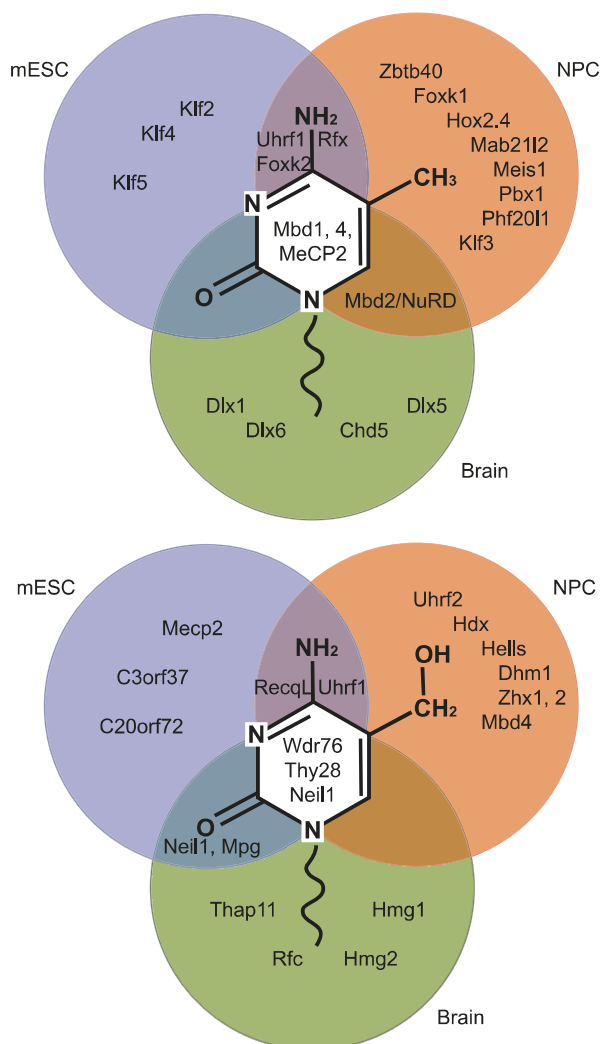
differentiated to NPCs in N2B27 medium and cultured in NSA medium, consisting of NSA MEM, 1% glutamine,  $1 \times$  N2 supplement, 10 ng/ml bFGF, and 10 ng/ml EGF.

### DNA Pull-Downs

Nuclear extracts were generated as described previously (Eberl et al., 2013; Vermeulen et al., 2010). DNA (see Table S4) immobilized on Dynabeads MyOne C1 was incubated with nuclear extract in 50 mM Tris-HCl (pH 8), 150 mM NaCl, 1 mM DTT, 0.25% NP40, and complete protease inhibitors (Roche, EDTA-free) in the presence of poly-dAdT. After extensive washes (using incubation buffer without poly-dAdT), bound proteins were in-gel digested using trypsin. After sample preparation, peptides were desalted on Stage-tips (Rappsilber et al., 2003).

### Mass Spectrometry

Peptides were separated using an EASY-nLC (Proxeon) connected online to an LTQ-Orbitrap Velos mass spectrometer (Thermo) as described (Smits et al., 2013). Raw data were analyzed using MaxQuant version 1.2.2.5 and searched against protein database ipi.MOUSE.v3.68.fasta. Using Perseus, data were filtered and scatter plots were made using R. The raw mass spectrometry data have been deposited to the ProteomeXchange Consortium



**Figure 7. Visualization of Cell-Type-Specific and General mC and hmC Readers**

Venn diagram showing examples of mC (A) and hmC (B) readers that were identified in mESCs (blue), NPCs (orange), and adult mouse brain (green). See also Figure S3 and Table S3.

(<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (Vizcaino et al., 2013) with the dataset identifier PXD000143.

#### Recombinant Protein Expression and DNA Pull-Downs

DNA-binding domains were cloned into the GST-containing PRP256NB vector. The Uhrf2(aa416–626) GST-fusion construct was kindly provided by Dr. Jiemin Wong. Protein expression was performed in *E. coli* BL21 codon<sup>+</sup> cells. Bacterial lysate was cleared by ultracentrifugation. DNA pull-downs were performed as described above with the addition of 10  $\mu$ M ZnCl<sub>2</sub> to the incubation buffer.

#### iBAQ

iBAQ was performed essentially as described in Schwanhäusser et al. (2011). A more detailed description can be found in the Extended Experimental Procedures.

#### LC-MS/MS Analysis of Genomic DNA

Cotransfections were performed in HEK293T cells, and genomic DNA was purified according to Münzel et al. (2010). Quantification of DNA nucleosides from genomic DNA is based on a further development of our isotope dilution method (Pfaffeneder et al., 2011 and data not shown). LC-MS/MS analysis was performed on an Agilent 6490 triple quadrupole mass spectrometer coupled to an Agilent 1290 UHPLC system. For general source- and compound-dependent parameters, see the Extended Experimental Procedures and Tables S5 and S6. The transitions of the nucleosides were analyzed in the positive-ion-selected reaction monitoring mode (SRM) operating MS1 and MS2 under unit mass resolution conditions.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2013.02.004>.

#### ACKNOWLEDGMENTS

We would like to thank Dr. Jiemin Wong for providing constructs. We are very thankful to Dr. Nikolay Outchkourov for providing NPCs, and we thank Tony Hyman and Ina Poser for providing the BAC-GFP WDR76 cell line. We would also like to thank Deepani Poramba Liyanage for support. Furthermore, we would like to thank all group members that contributed to fruitful discussions and Susan Kloet for critical reading of the manuscript. We thank the PRIDE team for their help uploading the raw mass spec data. Work in the Vermeulen lab is supported by a grant from the Netherlands Organization for Scientific Research (NWO-VIDI) and an ERC Starting Grant. The Carell group thanks the DFG Normalverfahren CA275/8-4, SFB 749, SFB 646, and the Volkswagen Foundation for financial support. Work in the Leonhardt group is supported by the SFB 646 and SPP 1463 grants from The Deutsche Forschungsgemeinschaft (DFG). M. Münzel and T.P. thank the Fonds der Chemischen Industrie for predoctoral fellowships. F.K. is supported through the International Research Support Initiative Program from the Higher Education Commission, Pakistan. H.v.l. is supported by a Veni fellowship from the Dutch Science Foundation Chemical Sciences, NWO-CW. K.L. and J.W. thank the DFG (WA1029/6) for support through SPP1356.

Received: October 9, 2012

Revised: January 11, 2013

Accepted: February 5, 2013

Published: February 21, 2013

#### REFERENCES

- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* 455, 818–821.
- Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y.J., Duan, S.L., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 455, 822–825.
- Bartels, S.J.J., Spruijt, C.G., Brinkman, A.B., Jansen, P.W.T.C., Vermeulen, M., and Stunnenberg, H.G. (2011). A SILAC-based screen for Methyl-CpG binding proteins identifies RBP-J as a DNA methylation and sequence-specific binding protein. *PLoS ONE* 6, e25884.
- Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* 143, 470–484.
- Bertonati, C., Punta, M., Fischer, M., Yachdav, G., Forouhar, F., Zhou, W.H., Kuzin, A.P., Seetharaman, J., Abashidze, M., Ramelot, T.A., et al. (2009). Structural genomics reveals EVE as a new ASCH/PUA-related domain. *Proteins* 75, 760–773.

- Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. *Mol. Cell* 38, 179–190.
- Bostick, M., Kim, J.K., Estève, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760–1764.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W.W., Jiang, J.M., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106–1117.
- Chowdhury, R., Yeoh, K.K., Tian, Y.M., Hillringhaus, L., Bagg, E.A., Rose, N.R., Leung, I.K.H., Li, X.S., Woon, E.C.Y., Yang, M., et al. (2011). The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* 12, 463–469.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.
- Defossez, P.A., and Stancheva, I. (2011). Biological functions of methyl-CpG-binding proteins. *Prog. Mol. Biol. Transl. Sci.* 101, 377–398.
- Dejosez, M., Krumenacker, J.S., Zitir, L.J., Passeri, M., Chu, L.F., Zhou, S.Y., Thomson, J.A., and Zwaka, T.P. (2008). Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell* 133, 1162–1174.
- Dennis, K., Fan, T., Geiman, T., Yan, Q.S., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.* 15, 2940–2944.
- Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. *Mol. Cell* 49, 368–378, Published online November 29, 2012. <http://dx.doi.org/10.1016/j.molcel.2012.10.026>.
- Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18, 553–567.
- Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M.C., Antes, I., and Leonhardt, H. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* 6, e21306.
- Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403, 916–921.
- Globisch, D., Münzel, M., Müller, M., Michalak, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* 5, e15367.
- Guo, J.U., Su, Y.J., Zhong, C., Ming, G.L., and Song, H.J. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145, 423–434.
- Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* 329, 78–82.
- Hashimoto, H., Liu, Y.W., Upadhyay, A.K., Chang, Y.Q., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X.D. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* 40, 4841–4849.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q.Y., Ding, J.P., Jia, Y.Y., Chen, Z.C., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307.
- Hubner, N.C., and Mann, M. (2011). Extracting gene function from protein-protein interactions using Quantitative BAC Interactomics (QUBIC). *Methods* 53, 453–459.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303.
- Jones, D.L., Howard, M.A., Stanco, A., Rubenstein, J.L.R., and Baraban, S.C. (2011). Deletion of Dlx1 results in reduced glutamatergic input to hippocampal interneurons. *J. Neurophysiol.* 105, 1984–1991.
- Kastan, M.B., Onyekwere, O., Sidransky, D., Vogelstein, B., and Craig, R.W. (1991). Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* 51, 6304–6311.
- Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468, 839–843.
- Konstandin, N., Bultmann, S., Szewagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P.M., Schneider, S., et al. (2011). Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia* 25, 1649–1652.
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Lee, D.W., Zhang, K.J., Ning, Z.Q., Raabe, E.H., Tintner, S., Wieland, R., Wilkins, B.J., Kim, J.M., Blough, R.I., and Arcaci, R.J. (2000). Proliferation-associated SNF2-like gene (PASG): a SNF2 family member altered in leukemia. *Cancer Res.* 60, 3612–3622.
- Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286, 35334–35338.
- Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell* 151, 1417–1430.
- Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmler, V., Michalak, S., Müller, M., Biel, M., and Carell, T. (2010). Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* 49, 5375–5377.
- Myant, K., Termanis, A., Sundaram, A.Y.M., Boe, T., Li, C., Merusi, C., Burrage, J., de Las Heras, J.I., and Stancheva, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Res.* 21, 83–94.
- Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C.A., Hagemeyer, C., and Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* 50, 7008–7012.
- Pichler, G., Wolf, P., Schmidt, C.S., Meilinger, D., Schneider, K., Frauer, C., Fellingner, K., Rottach, A., and Leonhardt, H. (2011). Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J. Cell. Biochem.* 112, 2585–2593.
- Potts, R.C., Zhang, P.S., Wurster, A.L., Precht, P., Mughal, M.R., Wood, W.H., 3rd, Zhang, Y.Q., Becker, K.G., Mattson, M.P., and Pazin, M.J. (2011). CHD5, a brain-specific paralog of Mi2 chromatin remodeling enzymes, regulates expression of neuronal genes. *PLoS ONE* 6, e24515.
- Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* 75, 663–670.
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342.
- Sengupta, P.K., Ehrlich, M., and Smith, B.D. (1999). A methylation-responsive MBDP/RFX site is in the first exon of the collagen alpha2(I) promoter. *J. Biol. Chem.* 274, 36649–36655.
- Smits, A.H., Jansen, P.W., Poser, I., Hyman, A.A., and Vermeulen, M. (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res.* 41, e28.
- Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., and Leonhardt, H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* 38, e181.

- Tahiliani, M., Koh, K.P., Shen, Y.H., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.
- Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464, 1082–1086.
- Toyota, H., Jiang, X.Z., Asakura, H., and Mizuguchi, J. (2012). Thy28 partially prevents apoptosis induction following engagement of membrane immunoglobulin in WEHI-231 B lymphoma cells. *Cell. Mol. Biol. Lett.* 17, 36–48.
- Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., and Mann, M. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 142, 967–980.
- Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., et al. (2013). The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 41(Database issue), D1063–D1069.
- Wang, B., Lufkin, T., and Rubenstein, J.L.R. (2011). Dlx6 regulates molecular properties of the striatum and central nucleus of the amygdala. *J. Comp. Neurol.* 519, 2320–2334.
- Wossidlo, M., Arand, J., Sebastiano, V., Lepikhov, K., Boiani, M., Reinhardt, R., Schöler, H., and Walter, J. (2010). Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J.* 29, 1877–1888.
- Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Wang, P., Xiao, M.T., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of  $\alpha$ -ketoglutarate-dependent dioxygenases. *Cancer Cell* 19, 17–30.
- Yildirim, O., Li, R.W., Hung, J.H., Chen, P.B., Dong, X.J., Ee, L.S., Weng, Z.P., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147, 1498–1510.
- Zhang, J.Q., Gao, Q.Q., Li, P.S., Liu, X.L., Jia, Y.H., Wu, W.C., Li, J.W., Dong, S., Koseki, H., and Wong, J.M. (2011). S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell Res.* 21, 1723–1739.



# Supplemental Information

## EXTENDED EXPERIMENTAL PROCEDURES

### SILAC Labeling of ES Cells

IB10 murine Embryonic stem cells were cultured feeder-free on gelatin coated dishes in medium consisting of 500 ml SILAC Dulbecco's Modified Eagle Medium without arginine, lysine and glutamine (PAA, E15-086), supplemented with 15% MESC serum substitute (Thermo Scientific), Glutamine, Penicillin/Streptomycin, 1x Non-essential amino acids, sodium pyruvate, 73 µg/ml L-Lysine (light/K<sup>0</sup> (Sigma, A6969), medium/K<sup>4</sup> (Sigma, 616192 or Silantes, 211103912) or heavy/K<sup>8</sup> (Sigma, 608041 or Silantes, 211603902)) and 29.4 µg/ml arginine (light/R<sup>0</sup> (Sigma, A6969), medium/R<sup>6</sup> (Sigma, 643440 or Silantes, 201203902) or heavy/R<sup>10</sup> (Sigma, 608033 or Silantes, 201603902)), LIF (1000 U/ml), β-mercaptoethanol and 2i compounds (CHIR99021 and PD0325901, 3 and 1 µM respectively). Cells were cultured in SILAC medium until labeling efficiency exceeded 95% after which cells were expanded and harvested to generate nuclear extracts.

### NPC Culturing

Neuronal progenitor cells were kindly provided by Dr. N. S. Outchkourov. They were cultured in medium consisting of NSA MEM (Euromed EVM0883LD), 1% glutamine, 1x N2 supplement, 10 ng/mL bFGF (RD systems 233-F3) and 10 ng/mL EGF (235-E9) on gelatin-coated dishes. Cells were detached from culture plates using accutase. Nuclear extracts were made as described below.

### Mice Brain Nuclear Extracts

Nuclei from adult mouse brain were purified by centrifugation through a sucrose cushion following homogenization, modified from (Lavery and Schibler, 1993). Then nuclei were lysed as described below.

### Nuclear Extract Preparation

This protocol is based essentially on Dignam et al. (Dignam et al., 1983). Briefly, cells were trypsinized and washed two times with PBS. Using a hypotonic buffer, the cells were swollen, after which the cells were lysed by dounce homogenizing in the presence of 0.15% NP40 and complete protease inhibitors. After centrifugation, the pellet consisting of nuclei was lysed by 90 min incubation in 2 volumes of nuclear lysis buffer (420 mM NaCl, 20 mM HEPES pH 7.9, 20% v/v glycerol, 2 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.1% NP40, complete protease inhibitor w/o EDTA (Roche) and 0.5 mM DTT). After centrifugation, the supernatant containing the soluble nuclear extract was aliquoted and snap frozen until further usage. Protein concentrations of the nuclear extracts were determined using the Biorad Protein assay.

### DNA Synthesis

The synthesis of the oligonucleotides was performed on an ABI 394 DNA/RNA Synthesizer (Applied Biosystems) using typical reagent concentrations (activator: 0.25 M benzylthiotetrazole in MeCN (10 ppm H<sub>2</sub>O), detritylation: 3% dichloroacetic acid in CH<sub>2</sub>Cl<sub>2</sub>, oxidation: 25 mM I<sub>2</sub> in MeCN/H<sub>2</sub>O/2,6-lutidine (11/5/1), capping: Ac<sub>2</sub>O/2,6-lutidine/MeCN (30 ppm H<sub>2</sub>O) (20/30/50) and 20% *N*-methylimidazole in MeCN (10 ppm H<sub>2</sub>O). The oligonucleotide syntheses were performed on 200 nmol low-volume polystyrene carriers using 0.1 M DNA CE-phosphoramidites: A (Bz-dA), C (Bz-dC), G (iBu-dG), T, mC (Bz-mC) obtained from Glen Research or Link Technologies. hmC, fC and caC phosphoramidites were synthesized according to literature (Münzel et al., 2010b) and incorporated into DNA using the standard protocol. Benzylthiotetrazole was prepared according to literature (Welz and Muller, 2002). The coupling times for the modified bases were increased to 3 min to ensure maximum coupling efficiency.

The mC and the unmodified strands were treated with ethanolic ammonia for cleavage of the carrier and removal of the permanent protecting groups. hmC, fC and caC containing DNA was cleaved and deprotected using 0.4 M NaOH in MeOH/H<sub>2</sub>O 4:1 for 18 hr at room temperature. After addition of 600 µl triethylammonium acetate (1 M) and centrifugation, the supernatant was concentrated to 30% of the original volume in a speedvac. Analysis and purification was performed on a Waters HPLC system (Waters Alliance 2695 with PDA 2996, preparative HPLC: 1525EF with 2482 UV detector) with VP 250/10 Nucleosil 100-7 C 18 columns from Macherey Nage! using a gradient of 0.1 M triethylamine/acetic acid in water and 80% acetonitrile. The quality of the strands was determined by MALDI-MS. The forward and reverse oligos were combined and annealed in 10mM Tris pH8; 50 mM NaCl and 1 mM EDTA. Biotin-14-ATP was used to fill in the TT-overhang using Klenow exo-, followed by purification of the DNA on sephadex-G50 columns.

### DNA Pull-Downs

For each DNA pull-down, 10 µg of DNA (see Table S4) was immobilized on 75 µl of Dynabeads MyOne C1 (Invitrogen) by incubating for 1 hr at room temperature in a total volume of 350 µl of DNA binding buffer (1M NaCl, 10mM Tris-HCl pH8, 1mM EDTA pH 8 and 0.05% NP40. Coupling of the DNA to the beads was always verified by agarose gel electrophoresis. Beads containing immobilized DNA were then incubated with 400 µg of nuclear extract in a total volume of 600 µl of protein binding buffer (50mM Tris-HCl pH8, 150 mM NaCl, 1mM DTT, 0.25% NP40 and complete protease inhibitors (Roche, EDTA-free)) in the presence of 10 µg poly-dAdT for 2 hr at 4°C. Baits were then washed three times with 0.5 ml of protein binding buffer after which beads containing different DNA modifications and different SILAC labels were combined and loaded on 4%–12% NuPage gradient gels (Invitrogen) (for example, C-beads with light extract were combined with mC beads that were incubated with heavy extract; forward pull-down).

For the label-free analysis, three separate DNA pull-downs with every bait were performed and each of these was loaded on gel separately. For Western blot validation using endogenous antibodies, protein amounts were scaled down by a factor of four.

### In-Gel Digestion

Samples were analyzed on 4%–12% precast NuPage gels (Invitrogen) and subsequently stained using colloidal blue staining (Invitrogen). Each lane was cut into 8–12 gel slices and each of these slices was subjected to in-gel trypsin digestion overnight. Tryptic peptides were desalted on Stage-tips (Rappsilber et al., 2003).

### Mass Spectrometry

Peptides were separated on an EASY-nLC (Proxeon) connected online to an LTQ-Orbitrap-Velos mass spectrometer. Spectra were recorded in CID mode. A gradient of organic solvent (5%–30% acetonitrile) was applied (120 min) and the top 15 most abundant peptides were fragmented for MS/MS, using an exclusion list of 500 proteins for 45 s.

### Data Analysis

Raw data were analyzed using Maxquant version 1.2.2.5 and the integrated Andromeda search engine against protein database ipi.MOUSE.v3.68. Using Perseus, data was filtered for contaminants, reverse hits, number of peptides (>1) and unique peptides (>0). Ratios were logarithmized ( $\log_2$ ) and groups (consisting of forward and reverse) were defined. Proteins were filtered to have at least 2 valid values in one of the groups and missing values were imputed based on a normal distribution (width = 0.2 and shift = 0), after which Significance B was calculated (Benj.Hoch.FDR = 0.05). Scatterplots were made using R. Proteins were defined to be significant when both forward and reverse significance  $p < 0.05$  and minimal ratios were  $> 2$  in both experiments. The H/L ratios shown in Figure 2A–C were calculated using the formula  $(\log(\text{forward ratio}) - \log(\text{reverse ratio}))/2$ .

### Label-Free Quantification

LFQ values, based on the summed measured intensities of all tryptic peptides of a single protein, allows for comparing the relative abundance of a protein in different pull-downs. Changes in the LFQ intensity of a protein between pull-downs with different DNA modifications indicate preferential binding of that protein to one modification over another. Raw data were analyzed using Maxquant version 1.2.2.5 and protein database ipi.MOUSE.v3.68.fasta. Settings that were different from SILAC analyses were: multiplicity set at 1 and the options for 'label-free quantification' and 'match between runs' were selected. Using Perseus, data were filtered for contaminants, reverse hits, number of peptides (>1) and unique peptides (>0). LFQ intensities were logarithmized ( $\log_2$ ). After defining each triplicate as a group, proteins were filtered to have at least 3 values in a single group, assuming that when a protein binds specifically to one modification, it may only be identified in the three pull-downs with that modification. The missing values were imputed using a normal distribution (width = 0.3, shift = 1.8). Groups were defined and the significant outliers were calculated using ANOVA (FDR = 0.025,  $S_0 = 2$  for NPC and  $S_0 = 0$  for brain). Correlation based clustering was done in R for the ANOVA-outliers only, using LFQ-values which had been normalized by row-mean-subtraction.

### Purification of GFP-Fusion Proteins for EMSA

HEK293T cells were transfected with expression constructs encoding for GFP-Tdg or GFP-Dnmt1. 48 hr after transfection, cells were lysed 30 min on ice in Lysis-Buffer (50 mM  $\text{NaH}_2\text{PO}_4$ , 150 mM NaCl, 10 mM Imidazole, 0.5 mM EDTA, 0.5% Tween, 1 g/l DNaseI, 2 mM  $\text{MgCl}_2$ , 0.5 mM  $\text{CaCl}_2$ , 1 mM PMSF, 1x Protease-Inhibitor-Mix M (SERVA Electrophoresis GmbH)). The lysate was cleared by centrifugation (14 000 rpm, 10 min, 4°C) and incubation of the supernatant with equilibrated Ni-NTA beads (QIAGEN) in IP-buffer (50 mM  $\text{NaH}_2\text{PO}_4$ , 150 mM NaCl, 10 mM Imidazole, 0.5 mM EDTA, 0.05% Tween). After centrifugation (2200 rpm, 2 min) the supernatant was added to equilibrated GB-P-Ni-NTA beads (Chromotek) in IP-buffer and rotated for 2 hr at 4°C. After washing three times with Washing-Buffer (50 mM  $\text{NaH}_2\text{PO}_4$ , 300 mM NaCl, 10 mM Imidazole, 0.1% Tween), the GFP-fusion proteins were eluted with 50 mM  $\text{NaH}_2\text{PO}_4$ , 150 mM NaCl, 250 mM Imidazole, 0.05% Tween. The elution buffer was exchanged to 20 mM TrisHCl, pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT for EMSA reactions. The glycosylase activity of the purified Tdg was tested on T/G mismatch containing DNA (data not shown).

### Electrophoretic Mobility Shift Assays of Fluorescent DNA Oligonucleotides with GFP-Fusion Proteins

GFP-Tdg and GFP-Dnmt1 at decreasing concentrations (200 nM, 150 nM, 100 nM, 50 nM, 25 nM, 12.5 nM and 6.25 nM) were incubated for 30 min on ice with a 1:1 mixture of two distinctly labeled fluorescent 42mers (see Table S4, MWG-Eurofins, 250 nM each) containing a central CG site. The ATTO647N-labeled oligonucleotide contains only canonical bases whereas the ATTO550-labeled DNA bears different cytosine modifications (C, mC, hmC, fC and caC) or an abasic site at the CG position on both strands. Samples were run on a 6% non-denaturing polyacrylamide gel (pre-run 1 hr with 0.5x TBE) at 4°C. Oligonucleotide- and GFP-fluorescence was detected by the Typhoon Scanner (GE Healthcare). Quantifications were done with ImageJ.

### DNA Purification and Analysis after NE Incubation

DNA pull-downs were performed as described above, but all amounts were scaled up 3 times. As a control, all baits were also incubated in buffer plus poly-dAdT without nuclear extract for 2 hr at 4°C. The beads were washed 3x using 1 ml of incubation buffer

and 1x using 1M NaCl, 10mM Tris-HCl pH8, 1mM EDTA and 0.05% NP40, to reduce contamination with DNA from the nuclear extracts. Beads were then resuspended in 200  $\mu$ L incubation buffer and DNA was purified using phenol/chloroform extraction from the beads. The DNA-strands were finally dissolved in milliQ, enzymatically hydrolyzed to nucleosides and analyzed in triplicate (15 pmol each) by MALDI-MS or LC-MS/MS.

### GFP Pull-Downs

HeLa wild-type cells and a BAC-GFP transgenic cell line (WDR76) were cultured in SILAC medium for eight cell doublings, after which cells were expanded and nuclear extracts were made. For each pull-down 20  $\mu$ L of GFP-trap slurry (50% v/v; Chromotek) was washed and incubated for 90 min at 4°C with 1 mg of nuclear extract of (WT L, WT H, GFP L and GFP H) in a total volume of 400  $\mu$ L incubation buffer (300 mM NaCl, 20 mM HEPES KOH pH 7.9, 20% v/v glycerol, 2 mM  $MgCl_2$ , 0.2 mM EDTA, 0.1% NP40, complete protease inhibitor w/o EDTA (Roche) and 0.5 mM DTT) in the presence of 2  $\mu$ L ethidium bromide (10mg/ml, final concentration 50  $\mu$ g/ml). Beads were then washed two times with this incubation buffer, twice with PBS + 0.5% NP40 and two times with PBS only. During the last wash, beads of light control and heavy GFP pull-down were mixed and vice versa. Bound proteins were then subjected to on-bead trypsin digestion (Hubner and Mann, 2011) and significant proteins were determined as described for the SILAC DNA pull-downs. For the GFP-Tdg pull-down, mESC were cultured in normal mESC medium and a transient transfection with the GFP-Tdg plasmid (15  $\mu$ g/15cm dish) using PEI (ratio DNA:PEI = 1:3) was performed. GFP-Tdg was purified in a label-free method, thus 3 pull-downs were performed using GFP-trap beads and as a control the same extract was incubated in triplicate with control blocked agarose beads (Chromotek). For each pull-down 20  $\mu$ L of bead slurry (50% v/v) was washed and incubated for 90 min at 4°C with 1 mg of the nuclear extract in a total volume of 400  $\mu$ L incubation buffer (150 mM NaCl, 50 mM Tris-HCl pH8, 1mM DTT, 0.25% NP40 and complete protease inhibitor w/o EDTA (Roche)) to mimic the conditions of the DNA pull-downs as close as possible in the presence of 50  $\mu$ g/ml of ethidium bromide. Beads were then washed two times with 0.5 ml of incubation buffer, twice with PBS + 0.5% NP40 and two times with PBS only, after which bound proteins were on-bead digested. The Tdg-GFP purification was analyzed using a permutation-based t test (FDR = 0.05 & S0 = 3) to determine significant interactors.

### Recombinant Protein Expression/DNA Pull-Downs

Klf4(aa396-483), KDM2B(aa606-647), Cxxc5(aa234-293), MBD3(aa1-77) and Rfx5(aa85-173) were cloned into PRP256NB vector, containing a GST with a C-terminal multiple cloning site. Uhrf2 (aa416-626) GST fusion was kindly provided by Dr. Jiemin Wong. hMBD2b-GST was provided by Stefanie Bartels.

Protein expression was performed in *E. coli* BL21-DE3 Codon+ by growing them at 37°C until OD<sub>600</sub> of 0.5, after which expression was induced using 1mM IPTG and culturing for 3 additional hours at 25°C. Cells were lysed in 50 mM Tris-HCl pH 8.0/ 20% sucrose/ 1 mM EDTA/ 0.5 mM PMSF/ 1 mM DTT/ 1  $\mu$ g/ml aprotinin using lysozyme and Triton X-100 and repeated freeze-thawing. Bacterial debris was removed by ultracentrifugation.

DNA pull-downs were performed using 2.5  $\mu$ g DNA coupled to 16.75  $\mu$ L MyOne beads and 5  $\mu$ L of bacterial lysate/ nuclear extract in 250  $\mu$ L total volume (50mM Tris-HCl pH8, 150 mM NaCl, 1mM DTT, 0.25% NP40 and complete protease inhibitors (Roche, EDTA-free)) in the presence of 2.5  $\mu$ g polyAdT. After 3 times of washing with 0.5 ml of this buffer, beads were boiled in sample buffer. 5% of the input material and 100% of the bound material was loaded on gel for Western blot analyses.

### Western Blot

Gels were blotted onto nitrocellulose membranes. Blots were blocked using 5%-skimmed milk in TBST. Used antibodies are: Mouse $\alpha$ MBD3 (IBL, 3A3), Goat $\alpha$ MBD2 (Everest Biotech, EB07538), Rabbit $\alpha$ RBBP5 (Bethyl, BL766), Goat $\alpha$ Jun-C (SantaCruz), Rabbit $\alpha$ DNMT1 (Abcam, ab13537), Rabbit $\alpha$ Carf (Abcam, ab140519), Rabbit $\alpha$ GST (Santa Cruz, SC-138), Rabbit $\alpha$ GFP (home made), Donkey $\alpha$ mouseHRP and Donkey $\alpha$ RabbitHRP.

### NMR-Spectroscopy-Based Interaction Study of Rfx5 and mC DNA

The winged-helix (WH) domain of human Rfx5 (residues 85-173, plus 18 additional residues at the N-terminus) was expressed as a GST-fusion in BL21-DE3 Codon+ bacterial strains at 25°C in M9 minimal medium with  $^{15}NH_4Cl$  and/or  $^{13}C$ -glucose. The protein was purified by binding to a Glutathione agarose (GA) column (Sigma) and eluted with 50 mM reduced glutathione (Sigma). After thrombin digestion, Rfx5-WH was purified over a Sephadex-75 (HiLoad 16/60) column in buffer A (50mM KPi pH 7, 100 mM KCl, 5 mM DTT, 0.5 mM PMSF and protease inhibitors). NMR samples used for backbone assignment contained ca. 0.3 mM WH domain in 90/10%  $H_2O/D_2O$  in buffer A. NMR spectra (HNCACB, CBCACONH, HNCA, and HNCO) were recorded at 298K on a 600 or 750 MHz Bruker Avance II spectrometer, processed using the NMRPipe package (Delaglio et al., 1995), and analyzed using CcpNmr Analysis (Vranken et al., 2005). Backbone assignments were obtained for 90 out of 106 residues in the Rfx5-WH construct.

Interaction study with mC DNA was done using an 18bp DNA fragment (see Table S4; (Biolegio)) carrying a single mC on each strand. Annealed DNA oligos were lyophilized and dissolved in buffer A to a stock concentration of 620  $\mu$ M. The Rfx5-WH domain (103  $\mu$ M) was titrated with mC DNA, and after each addition (11 points in total) the  $^1H$ - $^{15}N$  HSQC spectrum of Rfx5-WH as recorded (298K / 600 MHz Bruker Avance II). Since the DNA sequence used is not palindromic, the two mC may be inequivalent in their capability to bind Rfx5. At high DNA:Rfx ratios, several peaks appear split in two in a roughly 1:1 ratio, suggesting that although the Rfx5-WH domain senses the distinct DNA sequence context of the two mC sites, it recognizes both with similar affinities



(data not shown). Although a few residues showed non-linear titration profiles, most peak displacements were linear. For further analysis, the binding sites were treated as being independent, resulting in an apparent dissociation constant for the Rfx5-WH – mC interaction.

Titration data were fitted using MatLAB scripts (MATLAB version 7.13.0, The MathWorks Inc., 2011) using the fast-exchange assumption for residues with observed chemical shift perturbations between 10 and 30 Hz (fast-exchange regime; 15 residues) in a global fit. The error bars for the observed peak position was set to 1.2 Hz. The overall reduced chi-square for the fit was 2.17. The error in the fitted  $K_D$  was estimated using 1000 MonteCarlo simulations resulting in an average of  $3.2 \pm 0.9 \mu\text{M}$ . The range of acceptable fits was examined using F-statistics from a grid search, resulting in 95% probability limits of  $10 \text{ nM} < K_D < 16 \mu\text{M}$ .

A homology model of Rfx5-WH domain was constructed on the basis of the DNA-bound crystal structure of the Rfx1 winged helix domain (PDB-id: 1DP7; 35% sequence identity) using the SwissModel server (Schwede et al., 2003). The model was validated against the predicted backbone dihedral angles from the observed backbone chemical shifts using TALOS+ (Shen et al., 2009). The model of mC bound to the putative binding pocket was constructed in PyMol by superimposing the mC DNA from the UHRF1-mDNA crystal structure (PDB-id 3CLZ) onto the Rfx1-bound DNA, such that the binding pocket and mC are aligned. To achieve a proper fit, the mC base was set to a *syn*-conformation. The side chains orientations of K110 and Y161 were adjusted manually to minimize clashes.

### In Silico Analysis of Klf4 ChIP-Seq Profile and Bisulfite Sequencing Data in mESCs Cells and NPCs

Klf4 binding data (ChIP-seq) was taken from (Chen et al., 2008) (GSM288354), and DNA methylation data (whole-genome bisulfite sequencing) was taken from (Stadler et al., 2012) (GSE30202). Annotated Klf4 peak centers (mESC) were extended with 50 bp on both sides to obtain 100-bp Klf4 binding regions. The mean CpG methylation of each 100-bp region was calculated for mESCs and NPCs and plotted as a scatterplot (Suppl. Figure S1B). For each quadrant of this scatterplot, the genomic distribution of the 100-bp Klf4 binding regions was calculated and plotted as a pie chart (Suppl. Figure S1C). Promoters were defined as  $\pm 1$  kb upstream and downstream from transcription start sites of the RefSeq mm9 annotation. The DNA sequences of the 100-bp Klf4 binding regions were used to search for the GGCGTG motif, and the CpG methylation within these motifs was calculated. The obtained distribution was plotted as a histogram (Suppl. Figure S1D). Analyses were done using Python, Perl and R.

### iBAQ

iBAQ was performed essentially as described in (Schwanhäusser et al., 2011). 3.3  $\mu\text{g}$  of UPS2 standard (Sigma) was added to 10  $\mu\text{g}$  of nuclear extract, which was digested using the FASP protocol (Wiśniewski et al., 2009). In addition, 100  $\mu\text{g}$  of NE was digested using FASP after which the peptides were separated into 8 fractions using SAX. Each of these samples was measured during a 4 hr gradient of LC-MS/MS. A linear fit was made for the known amounts of the UPS2 standard and the measured iBAQ intensities in the 10  $\mu\text{g}$  sample. Using this curve, iBAQ values of all other identified proteins in the 10  $\mu\text{g}$  sample were converted to amounts. A linear fit was again made using these amounts and the iBAQ values in the eight SAX fractions, which were used to extrapolate absolute protein amounts of all identified proteins in these samples.

### Cell Culture and Transfection Experiments

The mammalian GFP-Tet1cd expression vector was generated by PCR amplification of mouse (E14) cDNA encoding the catalytic domain of Tet1 (amino acids 1365 to 2057) and N-terminal GFP fusion. HEK293T cells were grown at  $37^\circ\text{C}$  and 5%  $\text{CO}_2$  in Dulbecco's Modified Eagle Medium (DMEM, *Invitrogen 41966-029*) supplemented with 10% fetal bovine serum and 1% Penicillin-Streptomycin. Cells were passaged at 80% confluency. All transfections were performed using the *jetPRIME* system (*PEQLAB Biotechnologie GmbH*) according to the manufacturer's instructions. HEK293T cells were seeded 24 hr prior to transfection at a density of  $3 \times 10^6$  cells per 75  $\text{cm}^2$ -flask and incubated in 10 ml of medium at  $37^\circ\text{C}$  and 5%  $\text{CO}_2$  for 24 hr. Cotransfection of GFP-Tet1cd plasmid (6  $\mu\text{g}$ ) either with mouse Uhrf2-GFP plasmid DNA (6  $\mu\text{g}$ ) (Pichler et al., 2011) or 6  $\mu\text{g}$  of pCMV6-Cdk5Rap1-v2 (*Origene RG216600*) as an unrelated control was carried out in a 75  $\text{cm}^2$  flask containing 10 ml of fresh medium. The transfection solution (500  $\mu\text{l}$  of *jetPRIME* buffer, 12  $\mu\text{g}$  of plasmid DNA and 24  $\mu\text{l}$  of *jetPRIME* reagent) was added to the medium and the cells were incubated at  $37^\circ\text{C}$  and 5%  $\text{CO}_2$  for 48 hr. After removal of the medium the cells were washed once with PBS and then lysed for DNA extraction according to (Münzel et al., 2010). The DNA was enzymatically digested to the nucleosides and subsequently analyzed by LC-ESI-MS/MS.

### LC-MS/MS Analysis of Genomic DNA and Synthetic DNA

The following LC-MS/MS method for the quantification of DNA-nucleosides is based on a further development of our precise and sensitive isotope dilution method ((Pfaffeneder et al., 2011) and manuscript in preparation). In the following we shortly summarize the parameters of the method. Genomic or synthetic DNA was enzymatically digested to the nucleoside level. A specific amount of internal standards with a stable isotope label were spiked to the digestion mixture for precise quantification. The following labeled nucleosides were used as internal standards: [ $^{15}\text{N}_2$ ]-dC, [ $\text{D}_3$ ]-mC, [ $\text{D}_2$ ,  $^{15}\text{N}_2$ ]-hmC, [ $^{15}\text{N}_2$ ]-fC, [ $^{15}\text{N}_2$ ]-caC and [ $\text{D}_3$ ]-dT. In case of genomic DNA the dC- or dG-content was determined by LC-UV-Detection.

LC-MS/MS analysis was performed on an Agilent 6490 triple quadrupole mass spectrometer coupled to an Agilent 1290 UHPLC system. The general source-dependent parameters were as follows: Gas Temp  $50^\circ\text{C}$ , Gas Flow 15 L/min, Nebulizer 30 psi, Sheath

Gas Heater 300°C, Sheath Gas Flow 11 L/min, Capillary Voltage 2500 V and Nozzle Voltage 500 V. For compound-dependent parameters used for genomic DNA see Table S5, for compound-dependent parameters used for synthetic DNA see Table S6. The transitions of the nucleosides were analyzed in the positive ion selected reaction monitoring mode (SRM) operating MS1 and MS2 under unit mass resolution conditions.

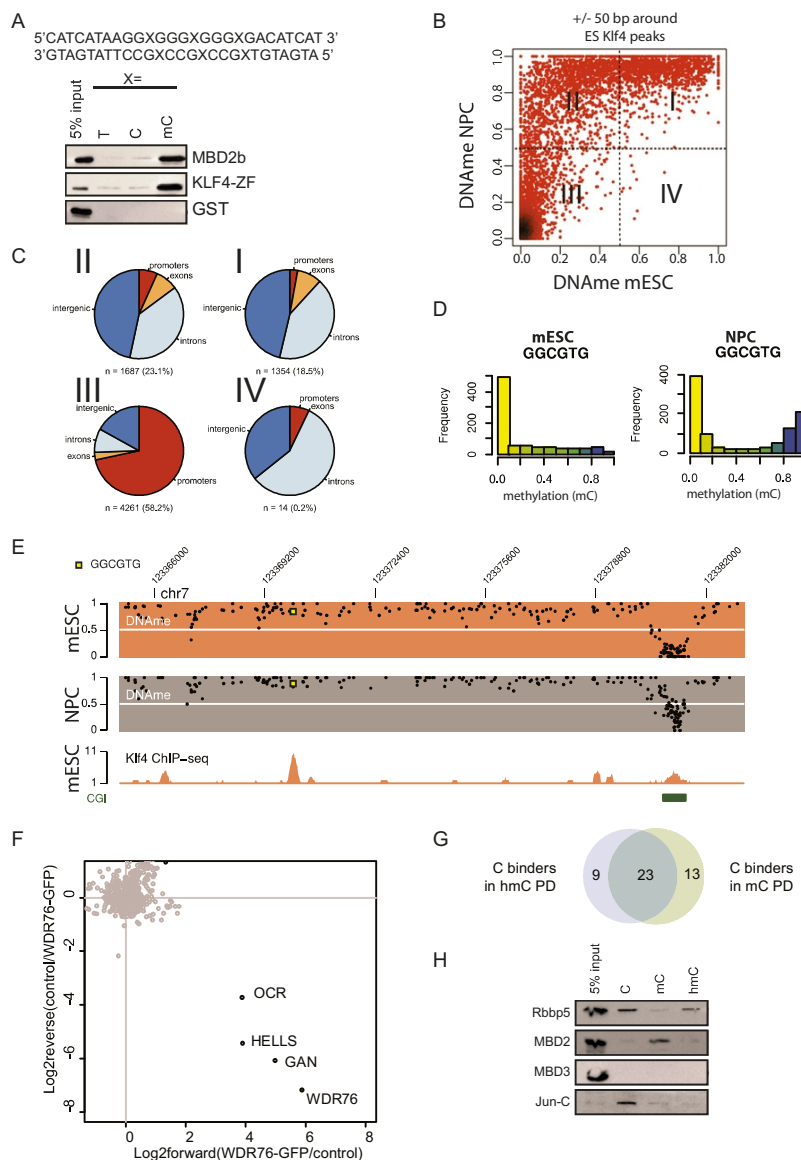
For the analysis a C8 column from Agilent was used (1.8  $\mu\text{m}$ , 2.1 mm x 150 mm). The compounds were separated by a gradient using water and acetonitril with 0.0075% formic acid. The column temperature was maintained at 30°C. The flow rate was 400  $\mu\text{L min}^{-1}$ , and the injection volume amounted to 29  $\mu\text{L}$ . The effluent up to 1.5 min (total run time of 12 min) was diverted to waste by a Valco valve in order to protect the mass spectrometer.

#### Validation of Quantification Method for Genomic DNA Modifications

In accordance with the FDA guidance for bioanalytical method validation, linearity, precision, and accuracy (i.e., recovery determined from spiked matrix samples) of the established method were investigated. Validation for the established LC-MS/MS quantification method was based on five different series (i.e., calibration functions and quality control samples) accomplished on different days. Calibration standards were analyzed at least in triplicates. Quality control samples to evaluate accuracy, intra- and inter-batch (see intra- and inter-assay) precision were determined using a biological sample with internal standards. Furthermore, each validation experiment was complemented by matrix blanks (analyzed in triplicates) to ensure selectivity and specificity of the method. Additionally, acceptable accuracy (80%–120%) as well as precision (<20% RSD) was required. Linear regression was applied to obtain calibration curves. Therefore, the peak area ratio ( $y$ ) of the unlabeled nucleoside to the internal standard versus the concentration ratio of the unlabeled nucleoside to the internal standard ( $x$ ) was plotted. Calibration functions were calculated without weighting. Long-term stability of aqueous solutions of the labeled and unlabeled nucleosides at a storage temperature of  $-20^{\circ}\text{C}$  was investigated over two months including several freeze and thaw cycles by analyzing the MS/MS-responses with each batch. Short-term stability at room temperature was studied in overnight experiments. In this process, the results of quantification by LC-ESI-MS/MS directly after preparing the samples were compared with those obtained from samples kept overnight at room temperature.

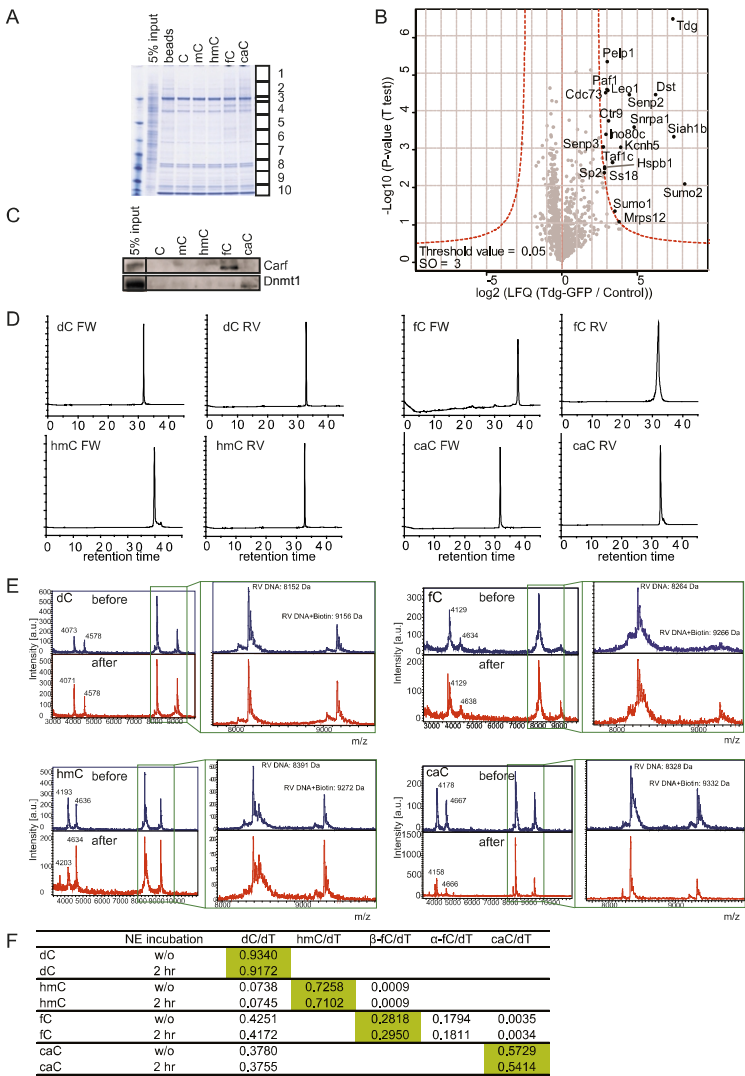
#### SUPPLEMENTAL REFERENCES

- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293.
- Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* 11, 1475–1489.
- Lavery, D.J., and Schibler, U. (1993). Circadian transcription of the cholesterol 7  $\alpha$  hydroxylase gene may involve the liver-enriched bZIP protein DBP. *Genes Dev.* 7, 1871–1884.
- Münzel, M., Globisch, D., Trindler, C., and Carell, T. (2010b). Efficient synthesis of 5-hydroxymethylcytosine containing DNA. *Org. Lett.* 12, 5671–5673.
- Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 31, 3381–3385.
- Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44, 213–223.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2012). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687–696.
- Welz, R., and Muller, S. (2002). 5-(benzylmercapto)-1H-tetrazole as activator for 2'-O-TBDMS phosphoramidite building blocks in RNA synthesis. *Tetrahedron Lett.* 43, 795–797.
- Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362.



**Figure S1. Genome-wide Localization of Klf4 Partially Correlates with DNA Methylation, Related to Figure 1**

(A) DNA pull-downs with recombinant GST-fusion proteins of DNA binding domains and Western blotting analysis.  
 (B) DNA methylation of Klf4 sites in ES cells and NPCs. Whole-genome bisulfite sequencing was used to determine DNA methylation within a window of +/- 50 bp around Klf4 peak centers. Darker coloring indicates high density of datapoints.  
 (C) Pie charts showing the genomic distribution of Klf4 sites as presented in the different quadrants of (B).  
 (D) Distribution of DNA methylation specifically within the GGCGTG sequence present underneath Klf4 sites.  
 (E) Example of DNA methylation profiles and Klf4 binding (ChIP-seq), showing binding of Klf4 to both methylated and unmethylated sites. Yellow squares indicate the presence of the GGCGTG sequence underneath Klf4 sites.  
 (F) SILAC-based GFP-purification from HeLa cells stably expressing WDR76-GFP. Significant interactors are indicated in black (high forward WDR76-GFP/control ratio, low reverse control/WDR76-GFP ratio).  
 (G) Venn diagram showing the overlap of C-specific readers in the mC and hmC DNA pull-downs from mESC nuclear extracts.  
 (H) Validation of C and mC specific binders by DNA pull-downs in HeLa nuclear extract and Western blotting for the endogenous proteins.



**Figure S2. Identification of TDG Interactors, Western Blot Verification of fC and caC Interactors, and Validation of Bait DNA Quality, Related to Figure 2**

(A) The indicated immobilized DNA baits were incubated with mouse nuclear extract. Following washes, bound proteins were analyzed by colloidal blue staining. Note that the elution profile of all these baits looks similar, indicating that specific interactors are masked by a large number of high abundant background binders.

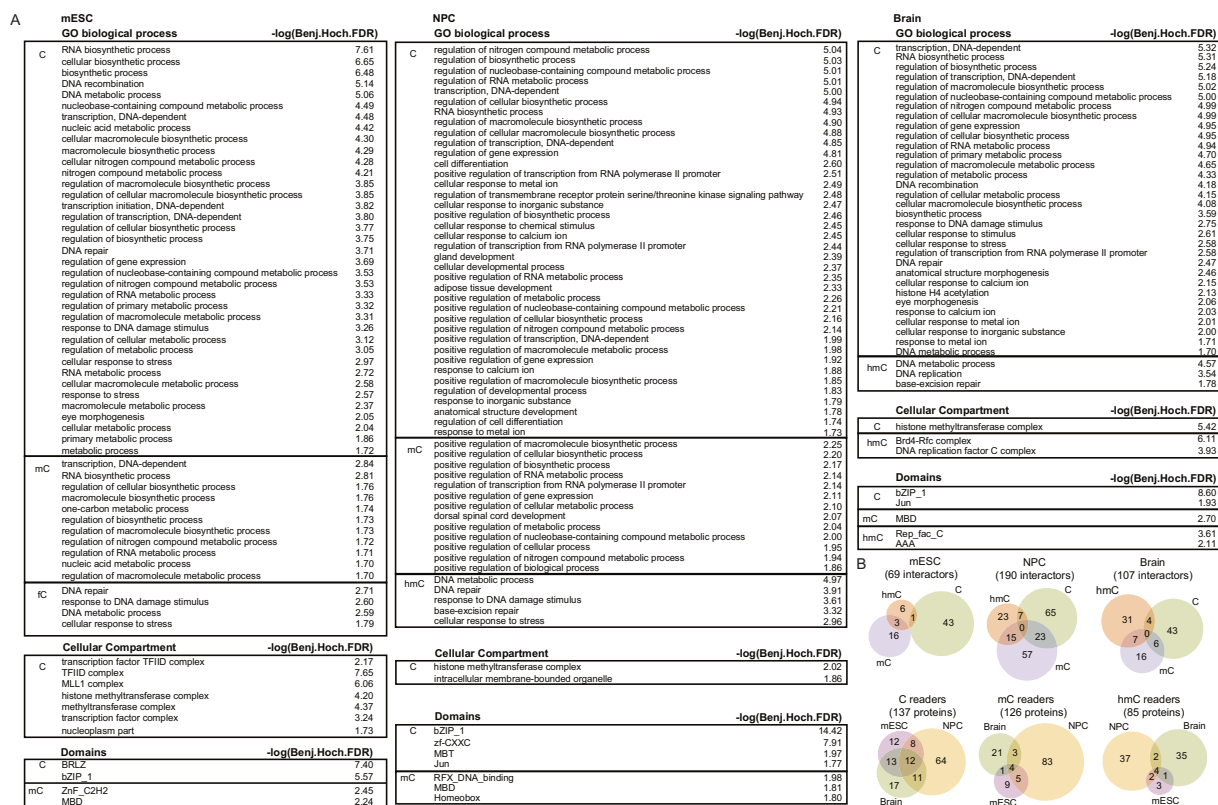
(B) Volcano plot of a label-free GFP-Tdg pull-down in mESC nuclear extract. Significant interactors of GFP-Tdg are identified by permutation-based t test (FDR = 0.05 & S0 = 3). The LFQ intensity of the GFP pull-down over the control is plotted against the  $-\log_{10}$  (p-value). The red line indicates the permutation-based FDR. Also see Table S1.

(C) Western blot validation of the fC-specific binding of Carf and caC-specific binding of Dnmt1 in mESC nuclear extract. A single empty lane was removed from the blot.

(D) HPL-Chromatograms of the purified FW and RV DNA obtained from solid phase DNA synthesis showing the purity of the employed strands.

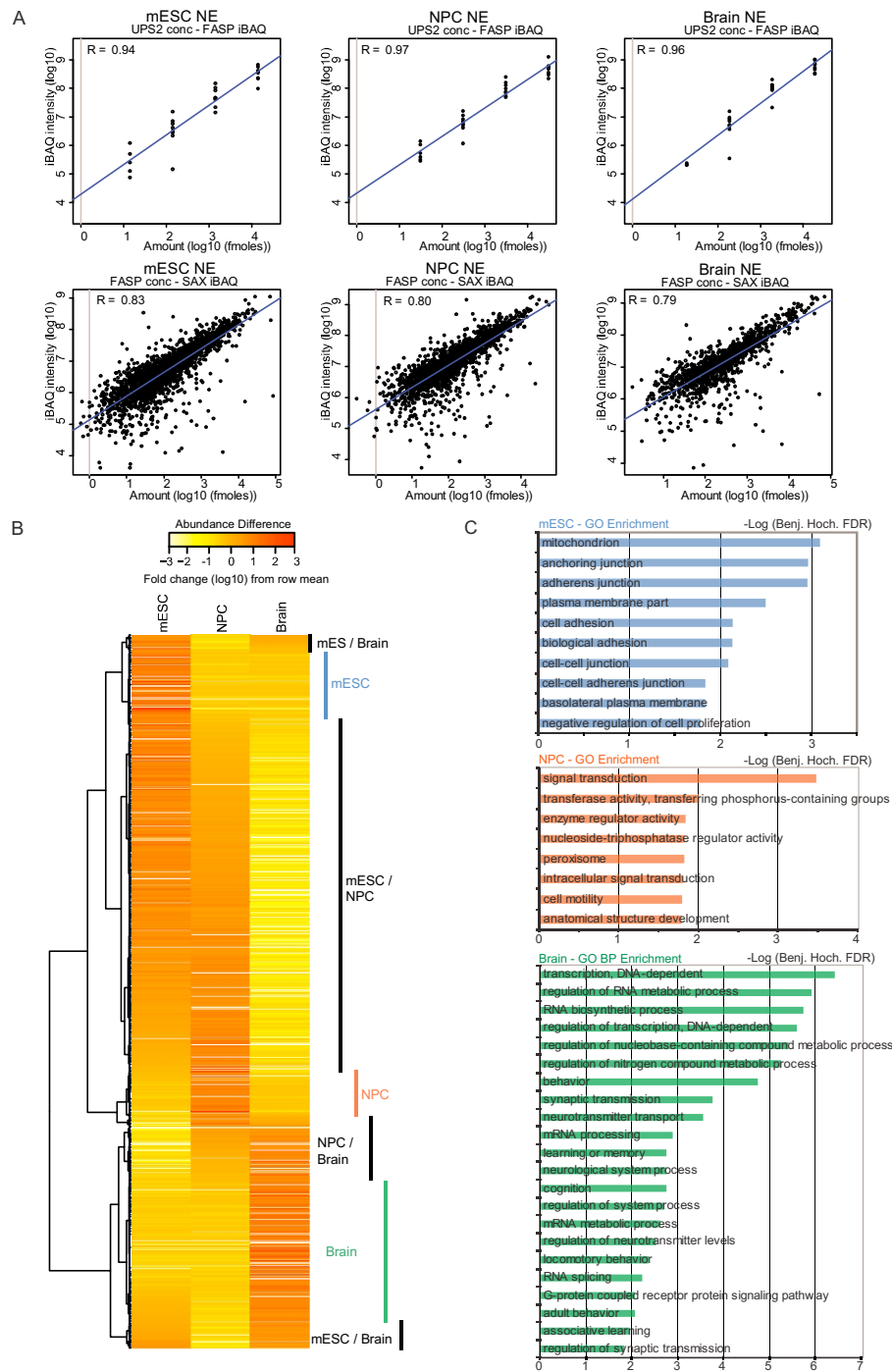
(E) The mass spectra of the DNA before (blue) and after (red) NE incubation as determined by MALDI MS showing the expected m/z before and after NE incubation. Major alterations of the DNA like degradation or strand breaks can be excluded.

(F) Synthetic DNA-strands which were used for DNA pull-downs were compared without (w/o) and with nuclear extract (NE) treatment (2 hr, 4°C) to proof the stability of the indicated modifications. The quantification of the nucleoside content was carried out by LC-MS/MS. For this, the DNA was digested to the nucleoside level and spiked with a specific amount of the following internal standards for precise quantification:  $[^{15}\text{N}_2]$ -dC,  $[\text{D}_2, ^{15}\text{N}_2]$ -hmC,  $[^{15}\text{N}_2]$ -fC,  $[^{15}\text{N}_2]$ -caC and  $[\text{D}_3]$ -dT. The absolute amount (pmol) of each nucleoside was calculated by calibration curves (not shown). Depicted are ratios of the modified nucleoside (pmol) to deoxy-thymidine (dT; pmol), which were obtained from three independent measurements. The relative standard deviation was between 0.3%–6.2%. No or only marginal loss of the modified nucleosides was observed.



**Figure S3. Modification and Cell-Type-Specific GO Term Enrichment Analysis, Related to Figures 1, 2, 3, and 4**

(A) Shows GO term enrichment and enriched domains for the different baits (C, mC, hmc, fC and caC) in mESC, NPCs and adult mouse brain. (B) Venn diagrams showing the overlap between C, mC and hmc readers within each cell type and the overlap between C, mC and hmc readers between mESCs, NPCs and adult mouse brain.



**Figure S4. iBAQ Analyses of mESC, NPC, and Adult Mouse Brain Nuclear Extracts, Related to Figure 5**

(A) Standard and linear regression curves for the iBAQ of protein abundance in the different nuclear extracts that were used for the DNA pull-downs.

(B) Correlation based clustering of proteins that show at least a 10-fold change in protein levels. Yellow is low abundance, red is high.

(C) GO term enrichment for mESC (indicated in blue in fig S4B), NPC (indicated in red in fig S4B) and adult mouse brain (indicated in green in fig S4) specific proteins.

## **2.2 TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation**

published June 19, 2014, in *Nucleic Acids Research*



# TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation

Udo Müller, Christina Bauer, Michael Siegl, Andrea Rottach and Heinrich Leonhardt\*

Department of Biology II, Ludwig-Maximilians University Munich and Center for Integrated Protein Science Munich (CIPSM), 82152 Planegg-Martinsried, Germany

Received March 28, 2014; Revised June 05, 2014; Accepted June 10, 2014

## ABSTRACT

The discovery of hydroxymethyl-, formyl- and carboxylcytosine, generated through oxidation of methylcytosine by TET dioxygenases, raised the question how these modifications contribute to epigenetic regulation. As they are subjected to complex regulation *in vivo*, we dissected links to gene expression with *in vitro* modified reporter constructs. We used an Oct4 promoter-driven reporter gene and demonstrated that *in vitro* methylation causes gene silencing while subsequent oxidation with purified catalytic domain of TET1 leads to gene reactivation. To identify proteins involved in this pathway we screened for TET interacting factors and identified TDG, PARP1, XRCC1 and LIG3 that are involved in base-excision repair. Knockout and rescue experiments demonstrated that gene reactivation depended on the glycosylase TDG, but not MBD4, while NEIL1, 2 and 3 could partially rescue the loss of TDG. These results clearly show that oxidation of methylcytosine by TET dioxygenases and subsequent removal by TDG or NEIL glycosylases and the BER pathway results in reactivation of epigenetically silenced genes.

## INTRODUCTION

DNA methylation at the C5-position of cytosine plays an essential role in a variety of fundamental processes, such as early embryonic development, X-chromosome inactivation, genome stability and imprinting (1,2). In vertebrates, this epigenetic modification is set by the three DNA methyltransferases DNMT1, DNMT3A and DNMT3B and the regulatory subunit DNMT3L (3–5).

Recently, it was discovered that the TET family of Fe(II)- and 2-oxoglutarate-dependent dioxygenases can successively convert 5-methylcytosine to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxylcytosine (caC) *in vitro* and *in vivo* (6–8). Three different TET proteins

(TET1, TET2 and TET3), each showing tissue-specific differential expression (9), have been identified in mouse and human (10). Functional studies indicate that they are involved in a variety of cellular processes including epigenetic reprogramming, differentiation, myelopoiesis and imprinting (11–13). Mutations of TET2 correlating with lower hmC levels and altered gene expression patterns have been linked to various hematopoietic malignancies (14,15).

The discovery of TET proteins and their catalytic products hmC, fC and caC has raised the question about the functions of these oxidized cytosine variants. They might serve as independent epigenetic signals and have been shown to recruit a distinct and dynamic set of ‘reader’ proteins in embryonic stem cells (ESCs) and differentiated cells (16). It also has been described that cytosine oxidation affects the efficiency of transcription by RNA polymerase II (17). However, the low abundance of fC and caC suggests that these cytosine variants are quickly processed *in vivo* and have been proposed to be intermediates in active DNA demethylation (18,19).

Whereas the mechanism of setting the methylation mark is well understood, the process of its removal has long been elusive. DNA demethylation may either occur by a passive process via the inhibition of DNMT1 maintenance methylation after replication (20–22) or by an active enzymatic reaction. In principle, there are three possibilities: first, the direct removal of the methyl group, second, the excision of either the methylated cytosine or third, of the entire nucleotide. It is currently proposed that the additional oxidized cytosine derivatives hmC, fC and caC are intermediates in active DNA demethylation, thereby contributing to epigenetic plasticity and transcriptional regulation (23,24).

Several biochemical studies revealed that thymine DNA glycosylase (TDG) can specifically bind to and excise fC and caC, resulting in abasic sites, which might be subsequently processed by the base-excision repair (BER) machinery (18,25). In general, the BER pathway repairs damaged DNA sites through recognition and excision of base lesions by substrate-specific glycosylases. The generated abasic site is subsequently cleaved by the AP endonuclease 1 (APEX1), leading to a single-strand break, which is recognized by PARP1 through its N-terminal zinc fingers. PARP1

\*To whom correspondence should be addressed: Tel: +49 89 2180 74229; Fax: +49 89 2180 74236; Email: h.leonhardt@lmu.de



then recruits XRCC1, LIG3 and DNA polymerase beta to complete the BER reactions (26–28). TDG depletion in mice causes embryonic lethality, and TDG deficient ESCs display prominent alterations of CpG modifications at a large number of gene regulatory regions (29,30).

Another discussed alternative for DNA demethylation is based on the initial deamination of hmC to hydroxymethyluracil (hmU) by members of the AID/APOBEC cytidine deaminase family (23). In the following step, hmU might be excised either by TDG, methyl-CpG-binding domain protein 4 (MBD4) or the single-strand-specific monofunctional uracil-DNA glycosylase 1 (SMUG1) (31–33). However, there is evidence that AID/APOBEC members are less active on modified cytosines *in vitro* or *in vivo*, challenging the prominence of the proposed deamination-linked demethylation pathway in living cells (34). Furthermore, a direct decarboxylation of caC to unmodified cytosine has been detected in ESC lysates, but no specific decarboxylase has been identified so far (35).

In addition to TDG, two members of the NEIL family of glycosylases (NEIL1 and NEIL3) have recently been identified as potential binders for oxidized cytosine derivatives (16). However, their function in TET-dependent demethylation has not been investigated to date.

To unravel the effects of TET-mediated cytosine oxidation on gene expression, we generated *in vitro* modified pOct4-reporter plasmids and monitored their *in vivo* expression in ESCs. Whereas methylation of the reporter DNA leads to silencing of gene expression, subsequent oxidation results in gene reactivation. We show that TET proteins interact with BER factors *in vivo* and propose that the observed oxidation-dependent gene reactivation requires the BER machinery. We demonstrate that initiation of this pathway is mainly dependent on TDG activity, but not on MBD4. Our results also indicate that the glycosylases NEIL1, NEIL2 and NEIL3 can contribute to an alternative BER pathway for DNA demethylation and cause gene reactivation.

## MATERIALS AND METHODS

### Cell culture and transfection

Human embryonic kidney 293T (HEK293T) and baby hamster kidney (BHK) cells containing a stably integrated lac operator array (36) were cultured in Dulbecco's modified Eagle's medium (Sigma) supplemented with 10% Fetal Calf Serum (FCS) (Biochrom) and 50 µg/µl gentamycin (PAA). HEK293T and BHK cells were transiently transfected using polyethylenimine pH 7.0 (Sigma) according to the manufacturer's instructions.

Mouse wild-type (wt) E14 as well as *Tdg*<sup>-/-</sup> and *Mbd4*<sup>-/-</sup> ESCs (29,37) were cultured on gelatin coated flasks or optical 96-well plates (Greiner) using 1000 U/ml LIF, 1 µM PD032591 and 3 µM CHIR99021 (Axon Medchem, (38)). ESCs were transfected with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions.

### *In vitro* methylation and oxidation of plasmid DNA

*In vitro* methylation of pOct4-GFP plasmid DNA was performed using M.SssI methyltransferase (New England Bi-

olabs) according to the manufacturer's instructions. The methylation status of the plasmid was tested by HpaII and MspI (Fermentas) digestion.

For the *in vitro* oxidation, GFP-TET1CD or GFP-TET1CD<sup>mut</sup> (H1652Y, D1654A) was purified from mammalian cells. In detail, HEK293T cells were transfected with an expression construct for GFP-TET1CD/TET1CD<sup>mut</sup> and immunoprecipitation was carried out using GBP-Ni-NTA beads. Proteins were eluted using imidazole. The *in vitro* methylated plasmid was diluted in TET reaction buffer (50 mM HEPES pH 8.0, 75 µM Fe(II), 2 mM Sodium-Ascorbate, 1 mM Di-Sodium-Ketoglutarate (39)) and added to the purified GFP-TET1CD.

### Digestion of hydroxymethylated plasmid with PvuRts1I

A total of 200 ng oxidized plasmid DNA and 100 ng of reference DNA fragments containing exclusively unmodified C, mC or hmC were digested with PvuRts1I (150 mM NaCl, 20 mM Tris pH 8.0, 5 mM MgCl<sub>2</sub>, 1 mM DTT) at 22°C for 20 min (40). The reaction was inactivated at 65°C for 10 min and digestion of the samples was analyzed by agarose-gelelectrophoresis.

### Co-immunoprecipitation (Co-IP) using the GFP-Trap

Note that 36 h after transfection, whole cell lysates of HEK293T cells were prepared using RIPA-lysis buffer (50 mM Tris pH 7.0, 150 mM NaCl, 0.1% sodium dodecyl sulphate (SDS), 0.5% sodium deoxycholate, 5 mM ethylenediaminetetraacetic acid (EDTA), 2.5 mM MgCl<sub>2</sub>, 0.5 mM CaCl<sub>2</sub>, 2 mM PMSF, 1x Mammalian Protease Inhibitor Cocktail and 1 µg/µl DNaseI). After centrifugation, 10% of the supernatant was collected as input fraction and the remaining supernatant was diluted in IP-buffer (10 mM Tris/Cl pH 7.5, 150 mM NaCl and 0.5 mM EDTA) to 800 µl. Green Fluorescent Protein (GFP)-Trap beads (Chromotek, (41)) pre-equilibrated with IP-buffer were added to the supernatant dilution and rotated for 1.5 h at 4°C. The GFP-beads were centrifuged and 10% of the supernatant was collected as flowthrough fraction. For western blotting, input, flowthrough and bead fractions were boiled with Laemmli buffer at 95°C for 10 min, loaded on an SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane (Biorad). Immunodetection was performed using mouse monoclonal anti-GFP (Roche, 11814460001) or rat monoclonal anti-RFP antibodies (42) and Alexa488/Alexa594 coupled secondary antibodies (Jackson ImmunoResearch).

For mass spectrometry analysis, protein samples were denatured with GdnHCl, precipitated with acetone and digested with trypsin. Peptide mixtures were analyzed using electrospray tandem mass spectrometry in collaboration with the Core Facility of the Max-Planck-Institute for Biochemistry, Martinsried. Experiments were performed with an LTQ Orbitrap mass spectrometer (Thermo Scientific). Spectra were analyzed with MaxQuant (43).

### Fluorescent-three-hybrid assay (F3H)

Transgenic BHK cells containing stably integrated lac-operator repeats (36) were grown to 60–70% confluence

on coverslips. The cells were transiently cotransfected with expression constructs encoding for LacI-GBP, murine RFP/mCherry- and GFP-fusion proteins (44). As controls the catalytically inactive mutants GFP-TET1<sup>H1652Y&D1654A</sup>, GFP-TET2<sup>H1304Y&D1306A</sup> and GFP-TET3<sup>H950Y&D952A</sup> were used. Note that 24 h after transfection cells were fixed with 4% formaldehyde in phosphate buffered saline (PBS), permeabilized with 0.5% Triton X-100/PBST, counterstained with 4',6-diamidino-2-phenylindole (DAPI) and mounted in Vectashield (Vector Laboratories). Imaging was performed using a Leica TCS SP5 confocal laser scanning microscope with a 63x/1.4 NA Plan-Apochromat oil immersion objective.

The Operetta automated imaging system (PerkinElmer) was used for F3H quantification (Harmony 3.6 software). After imaging, nuclei were detected based on DAPI signal. The lacO-spot was defined in the GFP channel and screened for enrichment at the RFP channel (intensity spot >1.2x mean intensity nucleus; see also Supplementary Figure S4).

#### High-throughput pOct4-reporter gene expression analysis

Wild type, *Tdg*<sup>-/-</sup> and *Mbd4*<sup>-/-</sup> ESCs were transiently transfected with unmodified, M.SssI methylated or *in vitro* oxidized pOct4-GFP plasmid DNA on coated optical 96-well plates (PerkinElmer). Note that 24 h after transfection, the cells were fixed with 4% formaldehyde/PBS, permeabilized with 0.5% Triton X-100/PBST and counterstained with DAPI.

Images were acquired using the Operetta automated imaging system with an 40x high NA objective and expression was quantified using Harmony 3.6 software (PerkinElmer). A total of 16 fields per well were imaged, cells were counted and segmented into nuclei and cytoplasm on the basis of DAPI and reporter mCherry/GFP signal. Prism software (GraphPad) was used for statistical analysis.

#### Generation of stable transgenic cell lines

*Tdg*<sup>-/-</sup> ESCs stably expressing GFP-fusions of wt TDG, TDG<sup>N151A</sup>, TDG<sup>N168D</sup> and TDG<sup>M280H</sup> were generated by transfecting the respective plasmids in the presence of the selection marker blasticidine followed by repeated sorting for GFP expression with the fluorescence-activated cell sorting (FACS) AriaII (Becton-Dickinson) system. Single cell sorting was used to generate clonal transgenic cell lines. GFP-expression of the single cell clones was analyzed using the Operetta system or western blotting.

#### Activity of GFP-TDG *in vitro*

GFP-TDG and the different mutants were expressed in HEK293T cells and immunoprecipitated as described above. Equal amounts of GFP-tagged protein immobilized on GFP-Trap beads were incubated with 0.4  $\mu$ M of DNA substrate in TDG reaction buffer (20 mM TrisHCl, pH 8.0 or pH 6.5, 100 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 1 mM DTT, 0.01 mM ZnCl<sub>2</sub>). In detail, these DNA substrates were fluorescently (ATTO550) labeled 42 bp oligonucleotides (GGA TGA TGA CTC TTC TGG

TCC GGA TGG TAG TTA AGT GTT GAG) (Eurofins MWG Operon) with a central modified CpG site: either hmCpG, fCpG or caCpG or harboring a G $\rightarrow$ T mismatch at this site. Incubation was performed in the presence of purified GFP-APEX1 for 2 h at room temperature. Following heat-inactivation of TDG (2 min, 95°C), fresh GFP-APEX1 was added and further incubated for 4 h at room temperature. An oligonucleotide containing a deoxyabasic site ('dSpacer', Eurofins MWG Operon) served as a positive control for APEX1 activity. DNA was analyzed on a denaturing 17% polyacrylamide gel with the Typhoon TRIO (GE Healthcare Life Sciences). Quantification was performed with ImageJ.

#### Genomic DNA extraction and slot blot analysis

Genomic DNA from ESCs was extracted using the Blood & Cell Culture Midi Kit (Qiagen) according to the manufacturer's instruction. Anti-oxidant BHT (200  $\mu$ M, Sigma) and deaminase inhibitor THU (200  $\mu$ M, Sigma) were added to the lysis buffer and elution buffer. The Bio-Rad slot blot system was used according to the manufacturer's instruction. Nitrocellulose membranes (Amersham) were crosslinked, blocked with 5% milk and immunostaining was performed using a mouse monoclonal antibody against mC (Eurogentec, 33D3) or rabbit polyclonal antibodies against hmC, fC and caC (Active motif: 39791, 61233, 61224). Alexa488-coupled secondary antibodies were used for detection and the membranes were scanned with the Typhoon TRIO (GE Healthcare Life Sciences). Quantification was performed with ImageJ.

#### Re-isolation of transfected plasmids

Note that 36 h after transfection, nuclei were extracted from the ESCs using the Blood & Cell Culture Midi Kit (Qiagen) according to the manufacturer's instructions. Plasmid DNA was re-isolated using the Qiaprep Plasmid Midi Kit (Qiagen). A total of 200 ng of re-isolated plasmid DNA was digested with 0.5  $\mu$ l HpaII (Fermentas).

## RESULTS

#### *In vitro* oxidation of Oct4 reporter plasmid DNA causes gene activation

Since the discovery of hmC, fC and caC, two major roles for these cytosine modifications have been proposed: first, as intermediates in active DNA demethylation, and second, as independent epigenetic marks. The latter has been investigated by identifying reader proteins for hmC, fC and caC in different murine tissues. These new DNA modifications recruit a dynamic set of readers including DNA repair factors and chromatin remodelers (16). TET-dependent cytosine oxidation has been shown to occur at a large number of gene regulatory elements and repetitive loci (30).

Here, we focus on the effects of hmC, fC and caC on gene expression. We transfected ESCs with Oct4 promoter-driven GFP and mCherry reporter plasmids (pOct4-GFP or pOct4-mCherry) in different modification states and analyzed expression of the reporter by quantitative imaging. To generate the different cytosine modifications *in*

*vitro*, the unmodified pOct4-reporter construct (CpG) was initially treated with the DNA-methyltransferase M.SssI, thereby creating fully methylated CpG sites (<sup>m</sup>CpG), and subsequently incubated with the purified catalytic domain of TET1 (TET1CD) to create oxidized cytosine residues (<sup>ox</sup>CpG; Figure 1a). Specific restriction enzymes were used to monitor the methylation and oxidation state of the plasmid DNA. The *in vitro* methylation of the reporter construct was confirmed by the mC-sensitive restriction endonuclease HpaII and the mC-insensitive enzyme MspI. Both MspI and HpaII fully cleave the unmethylated pOct4-reporter plasmid at CCGG sites, whereas the restriction pattern of the M.SssI-methylated reporter only appeared with MspI digestion (Figure 1b).

To follow the oxidation of mC to hmC by TET1CD, the reporter DNA was treated with the hmC-specific endonuclease PvuRtsII (40) (Supplementary Figure S1a). Treatment with TET1CD resulted in a gradual increase of hmC levels after 15, 45 and 90 min of incubation, visible as progressing fragmentation (Figure 1b). While using a catalytically inactive TET1CD mutant (TET1CD<sup>mut</sup>) as a control, no hmC levels were detected (<sup>ox</sup>\*CpG), indicating specific enzymatic mC oxidation by TET1CD (Supplementary Figure S1b).

Consistent with the restriction digest results, methylation and oxidation of the reporter plasmid could also be shown by slot blot analysis (Figure 1c). Incubation of methylated plasmid DNA with TET1CD resulted in an increase of not only hmC but also of fC and caC, demonstrating that purified TET1CD did carry out the three oxidation steps *in vitro* (Figure 1c and d). The mC signal decreased over time as hmC, fC and caC appeared in the presence of active TET1CD, while remaining constantly high with TET1CD<sup>mut</sup> (Supplementary Figure S1d).

Transfection of ESCs with the TET1CD-treated plasmid DNA allows to directly monitor the effect of the oxidized cytosines on gene expression, independent from *in vivo* TET activity. Reporter gene expression from either unmodified, *in vitro* methylated or oxidized pOct4-mCherry was visualized using confocal imaging or automated image acquisition for quantification. Transfection of ESCs with unmodified pOct4-mCherry resulted in a strong nuclear and cytoplasmic expression of the reporter, whereas expression drastically decreased when using the methylated construct. Interestingly, prominent reporter expression could be observed upon transfection of the oxidized plasmid DNA, but not with the <sup>ox</sup>\*CpG reporter DNA (Figure 1e). This suggests that reactivation of gene expression requires oxidation of methylcytosine by TET proteins.

### TET proteins interact with the BER machinery

Currently, three pathways for TET-mediated active DNA demethylation are discussed: TDG-dependent BER, deamination-dependent BER and direct decarboxylation of caC (18,23,35). Since the proteins responsible for gene reactivation in our assay might be physical interaction partners of the TET proteins, we performed an initial unbiased screen for interactors. We expressed GFP-TET1 in HEK293T cells, performed immunoprecipitation and analyzed the co-precipitated proteins by mass spectrometry.

Prominently, we found PARP1, XRCC1 and LIG3, a subset of proteins involved in the BER pathway (Supplementary Figure S2a). These results point toward the two BER-dependent demethylation mechanisms.

To further investigate the interplay between TET proteins and BER, we systematically performed Co-IP and a recently established F3H assay of all three TET proteins with the following BER factors: TDG, MBD4, SMUG1, NEIL1, NEIL2, NEIL3, PARP1, LIG3 and XRCC1. F3H allows to directly visualize protein–protein interactions in living cells (44). We therefore co-expressed full-length GFP-TET fusion proteins together with potential mCherry/RFP-tagged interactors (Supplementary Figure S3a) and LacI-GBP in BHK cells containing a stably integrated lac-operator array (45). The GFP-tagged bait is enriched at the lac-operator array via LacI-GBP and is visible as a single spot inside the nucleus. If the mCherry/RFP-tagged prey protein interacts with the bait, it colocalizes at the same spot (Figure 2a).

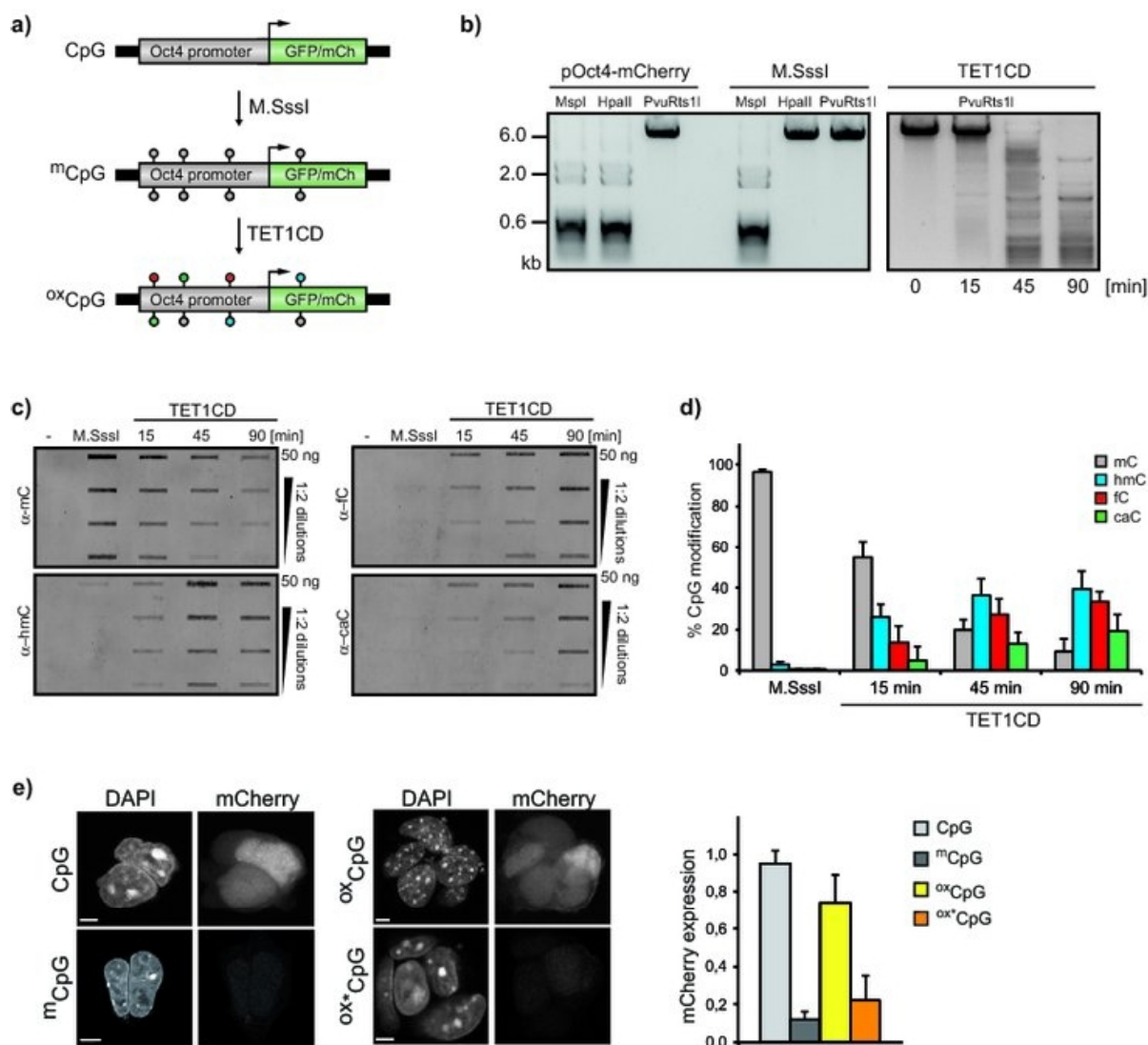
Consistent with the mass spectrometry results, interactions of TET1, TET2 and TET3 were observed with PARP1, LIG3 and XRCC1 both in Co-IP and F3H. Interestingly, several glycosylases also showed a clear interaction in both assays, among them TDG, MBD4, NEIL1, NEIL2 and NEIL3, but not SMUG1 (Figure 2b and c; Supplementary Figures S2b and S3b, c). Automated high-throughput image analysis was used to quantify the F3H results (Supplementary Figure S4). The interaction of all three TET proteins with TDG was the most robust and detectable in more than 80% of all analyzed cells. For the other factors, numbers vary between 40% and 75% (Figure 2d).

To exclude that the recruitment to the lacO-spot is dependent on the locally enriched cytosine oxidation products generated by TET proteins, we repeated the F3H quantification with catalytically inactive TET mutants (Figure 2e). The percentage of cells showing an interaction did hardly differ compared to the assay with active TET proteins. The only exception is XRCC1 where only half as many cells displayed a colocalization at the spot, suggesting a potential cooperative effect.

Taken together, these results suggest that TET proteins physically interact with the BER machinery and are therefore able to recruit these factors to the site of cytosine oxidation for immediate removal of the modified base *in vivo*.

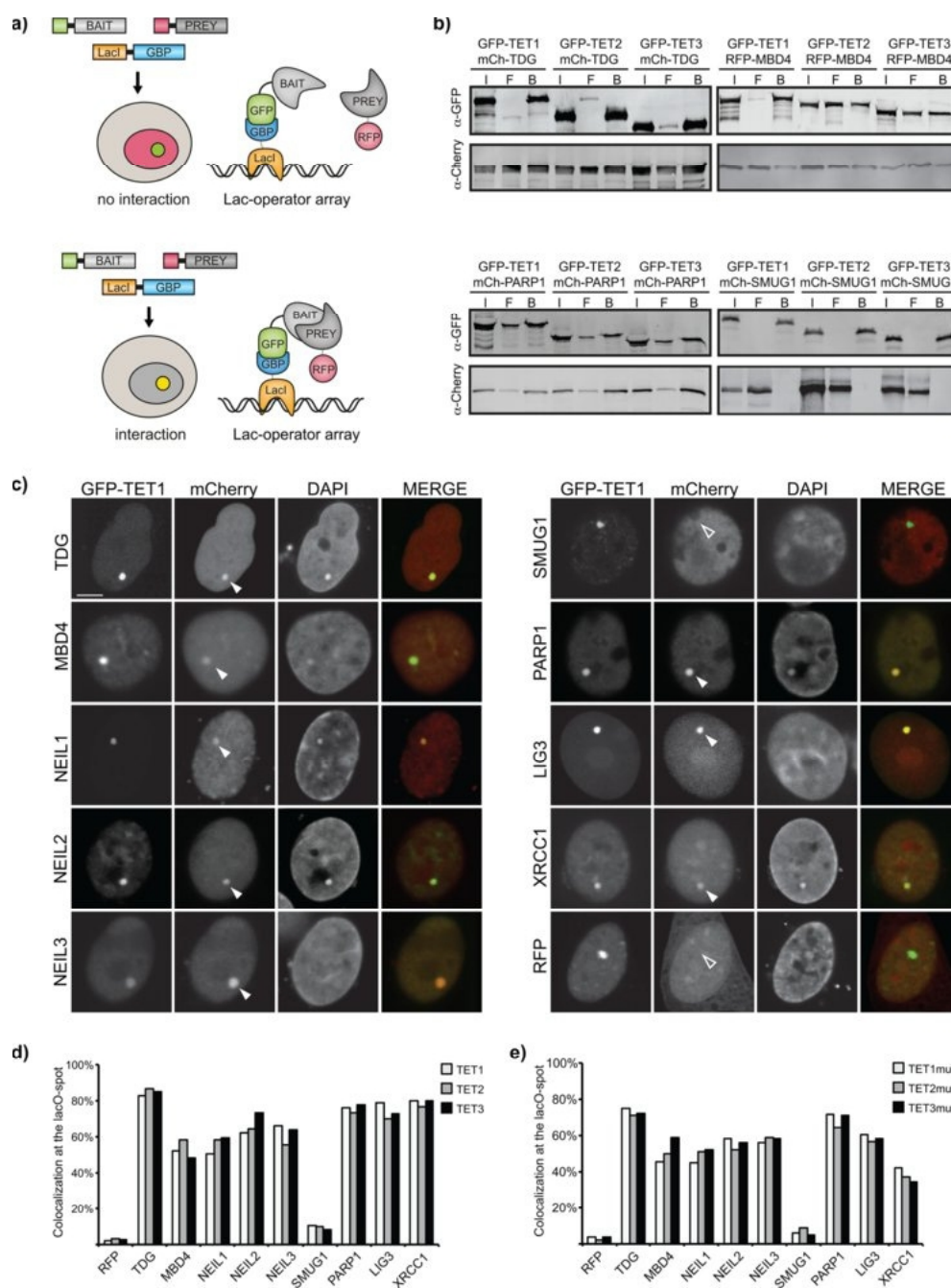
### TDG but not MBD4 mediates oxidation-dependent gene reactivation in ESCs

Since DNA glycosylases catalyze the first step of the BER pathway and are therefore the initiators of TET-dependent cytosine demethylation, we investigated their role in gene reactivation with the reporter gene assay. Besides TDG, also MBD4 has been implicated in DNA demethylation via excision of hmU, the deamination product of hmC (46). To investigate the role of these two glycosylases, we transfected *Tdg*<sup>-/-</sup> and *Mbd4*<sup>-/-</sup> ESCs (29,37) with either unmodified, *in vitro* methylated or oxidized pOct4-mCherry plasmid DNA. In contrast to wt E14 ESCs, *Tdg*<sup>-/-</sup> ESCs showed no reporter gene expression from the oxidized plasmid. However, *Mbd4* knockouts were able to fully reactivate gene expression from the oxidized reporter construct.



**Figure 1.** *In vitro* oxidation of mC causes gene reactivation in ESCs. (a) Schematic representation of *in vitro* reporter DNA modification: Unmethylated pOct4-reporter DNA was methylated using the CpG methyltransferase M.SssI. Incubation with purified TET1CD results in oxidation of mC sites to hmC, fC and caC. (b) M.SssI treatment of pOct4-mCherry results in full methylation as shown after restriction with the methylation sensitive enzyme HpaII. MspI cuts irrespective of the methylation state. The hmC-specific restriction endonuclease PvuRts1I detects increasing hmC levels during incubation of methylated pOct4-mCherry with TET1CD. (c) Cytosine modification states of untreated, methylated and TET1CD oxidized pOct4-mCherry plasmid DNA were detected by slot blot. A 2-fold serial dilution of the plasmid DNA was loaded and detected using antibodies against mC, hmC, fC and caC. A gradual increase of hmC, fC and caC signals was obtained with longer incubation time with TET1CD while the mC signal decreases accordingly. (d) Quantification of the slot blot signals of pOct4-mCherry after treatment with TET1CD shows increasing oxidation of mC to hmC, fC and caC. The sum of all CpG modification signals was set to 100%. Error bars indicate standard deviation ( $n = 3$ ). (e) ESCs were transfected with pOct4-mCherry plasmids containing either unmodified (CpG), methylated (mCpG), TET1CD-oxidized (oxCpG) or TET1CD<sup>mut</sup>-treated (ox\* CpG) cytosines. Confocal imaging and quantification show reporter gene silencing upon methylation and reactivation upon oxidation. Cells were fixed with formaldehyde and counterstained with DAPI. Scale bar: 5  $\mu$ m. (Right:  $n = 200,000$ ; error bars indicate standard deviation).





**Figure 2.** All three TET proteins interact with a variety of BER factors. (a) Scheme depicting the F3H assay for *in vivo* visualization of protein interactions: BHK cells containing a stably integrated lac-operator array were transfected with plasmids expressing a lac-repressor-GBP fusion protein, GFP-BAIT and mCherry/RFP-PREY. The Lac-repressor binds to the lac-operator array and recruits the GFP-BAIT through GBP. Proteins interacting with the BAIT are consequently enriched at the lac-operator array. (b) Co-IP with subsequent SDS-PAGE and western blotting shows interaction of the three TET proteins with TDG, PARP1, MBD4 but not SMUG1. GFP-tagged TET proteins and the respective mCherry-fusions were expressed in HEK293T cells and immunoprecipitated with the GFP-Trap (I: Input; F: Flowthrough; B: Bound). (c) F3H was used to confirm TET1 interactions with different factors involved in BER. GFP-TET1 is enriched at the lac-operator array and mCherry-tagged interacting factors are recruited to the same spot (solid triangle). SMUG1 shows no co-localization at the lacO-spot (empty triangle). Scale bar: 5  $\mu$ m. (d) Quantification of the F3H assay of all three TET proteins with the indicated BER factors. Bars represent percentage of cells that show colocalization at the lacO-spot ( $n > 200$ ). (e) As in (d) but with catalytically inactive TET mutants.

The rescue experiment with wt GFP-TDG re-established the ability of *Tdg*<sup>-/-</sup> to express mCherry from the <sup>ox</sup>CpG plasmid, whereas the transient rescue of *Mbd4*<sup>-/-</sup> cells with GFP-MBD4 led to no significant differences with regard to reporter gene expression (Figure 3a).

High-throughput image analysis of 200 000 cells revealed that the signal from the oxidized plasmid in wt E14 ESCs is about 80% of the signal from unmodified reporter DNA. In *Tdg*<sup>-/-</sup> cells, the <sup>ox</sup>CpG signal drops almost to the level of the fully methylated reporter. Thus, the <sup>ox</sup>CpG construct remains silent in *Tdg*<sup>-/-</sup> ESCs. The stable rescue with wt GFP-TDG led to a recovery of mCherry expression to about 70% of the signal from the unmodified reporter. Knockout of *Mbd4* and also the corresponding rescue did not alter expression levels compared to wt E14 ESCs (Figure 3b), indicating that oxidation-dependent gene reactivation requires TDG but not MBD4 expression.

To gain insight into the mechanism by which TDG mediates gene activation, we recovered the transfected reporter plasmid DNA from wt E14 ESCs, *Tdg*<sup>-/-</sup> and *Mbd4*<sup>-/-</sup> ESCs. We analyzed the modification status by digestion with HpaII, which specifically cuts unmodified cytosines in a CCGG context (Supplementary Figure S1c). We observed a reappearance of the HpaII restriction pattern on the <sup>ox</sup>CpG plasmids isolated from wt E14 and *Mbd4*<sup>-/-</sup> ESCs indicating that TET-dependent demethylation occurred *in vivo*. HpaII only displayed minor activity on <sup>ox</sup>CpG reporter DNA re-isolated from *Tdg*<sup>-/-</sup> cells (Figure 3c). This provides strong evidence that the substitution of oxidized cytosine by unmodified cytosine is the major mechanism for the observed gene reactivation and that this substitution depends on TDG.

### TDG activity is required for reporter gene reactivation

To investigate whether the glycosylase activity of TDG is responsible for the observed gene reactivation, we generated *Tdg*<sup>-/-</sup> rescue cell lines, which stably express GFP-fusions of either wt, catalytically inactive (N151A), DNA binding deficient (M280H) or caC-specific (N168D) TDG (31,47) at equal levels (Supplementary Figure S5a). The murine caC-specific mutant corresponds to a published human TDG mutant (48). Expression of the oxidized mCherry-reporter plasmid in the stable rescue with wt GFP-TDG increased almost to the levels of wt E14 ESCs in line with the results from the transient rescues. In contrast, the catalytically inactive mutant (N151A) or DNA binding deficient (M280H) TDG was not able to recover reporter expression. Cells stably expressing the caC-specific TDG (N168D) were only capable of partially restoring reporter expression (Figure 4a and b; Supplementary Figure S5b). This demonstrates that base excision by active TDG is essential for gene reactivation.

To further characterize the activity of wt TDG as well as the TDG mutants, we established an *in vitro* assay based on a defined DNA substrate with a single modification site. Base excision by TDG generates an abasic site, which can specifically be converted into a single-strand break by purified APEX1 and can be detected on a denaturing gel. Thus, this assay mimics the first two steps of the BER reactions and is also applicable on fluorescently labeled DNA sub-

strates, in contrast to the previously described ‘nicking assay’, in which alkaline treatment and subsequent boiling is used to create single-strand breaks (49,50).

TDG has long been known to repair G⋅T mismatches (51) and was recently found to excise fC and caC (18,25). We could show that wt TDG is more active on fC and caC than on G⋅T mismatches, its eponymous substrate. No activity was detected on hmC-containing DNA. Although the caC-specific TDG mutant (N168D) was able to partially reactivate reporter gene expression *in vivo*, we could only detect basal activity on the caC substrate *in vitro* (Figure 4c, Supplementary Figure S5d). Since the activity of TDG *in vitro* is pH-dependent (48), we repeated the assay at pH = 6.5. Under these conditions, the preference of the caC-specific TDG mutant (N168D) toward caC could be confirmed (Supplementary Figure S5d). Taken together, these data suggest that the excision of fC and caC by TDG is essential for TET-mediated demethylation and causes reactivation of gene expression.

### NEIL1, NEIL2 and NEIL3 glycosylases can partially compensate for loss of TDG

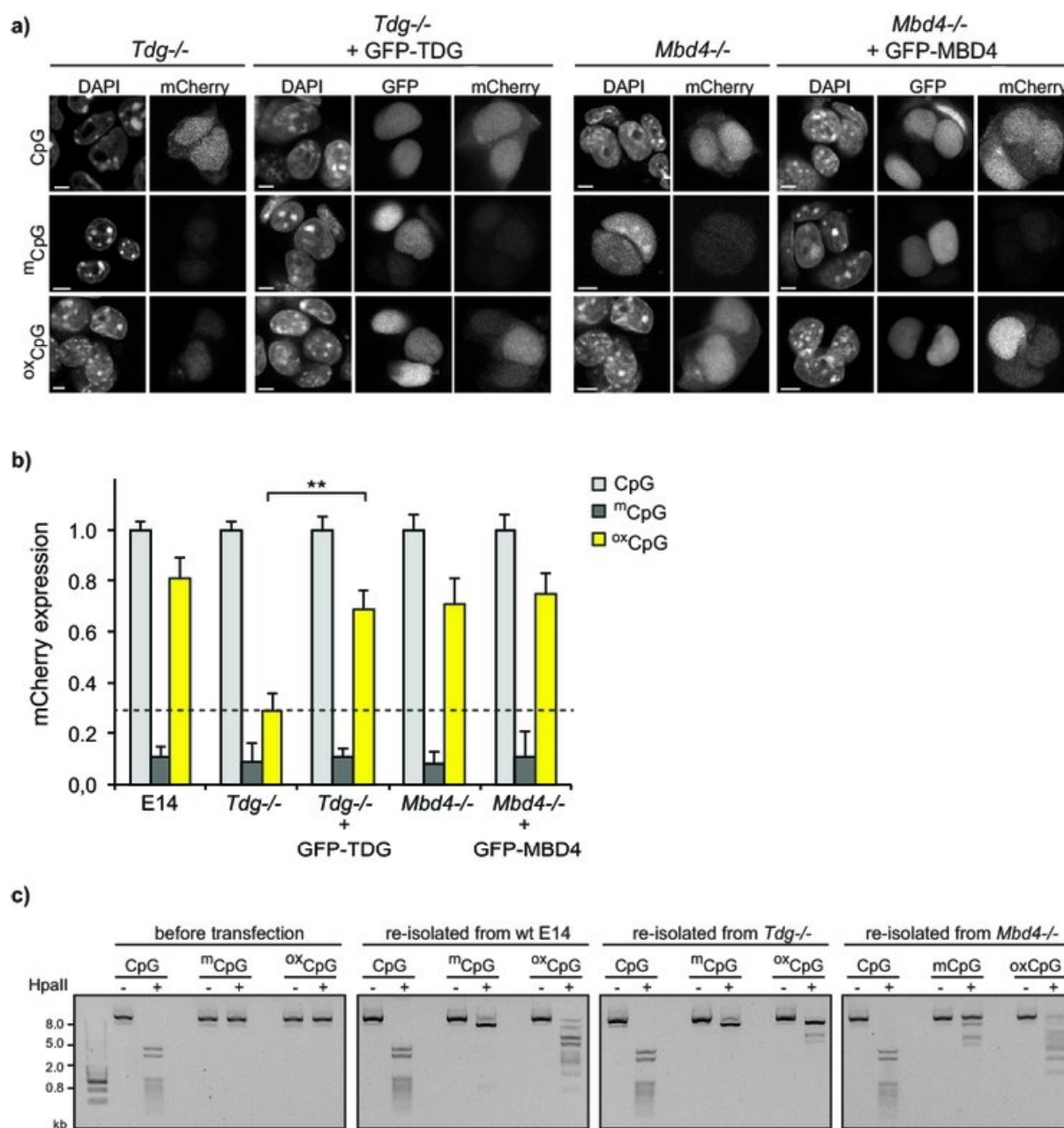
Besides TDG, we identified the family of NEIL glycosylases as interactors of TET proteins (Figure 2c). Interestingly, NEIL1 and NEIL3 have also been described as binders of hmC, fC or caC cytosines in ESCs in a proteome wide screen (16). However, their function in this context has not been investigated so far.

To elucidate whether the NEIL glycosylase family contributes to gene reactivation, we measured the expression of the modified reporter plasmids in *Tdg*<sup>-/-</sup> cells transiently overexpressing NEIL1, NEIL2 or NEIL3 at similar levels (Figure 5a). Interestingly, we observed a significant increase of pOct4-GFP expression in *Tdg*<sup>-/-</sup> ESCs rescued with mCherry-NEIL1, NEIL2 or NEIL3 in comparison to *Tdg*<sup>-/-</sup> ESCs. However, lower expression levels as in rescues with wt TDG were detected. RFP-MBD4 was not able to rescue the *Tdg*<sup>-/-</sup> phenotype (Figure 5b).

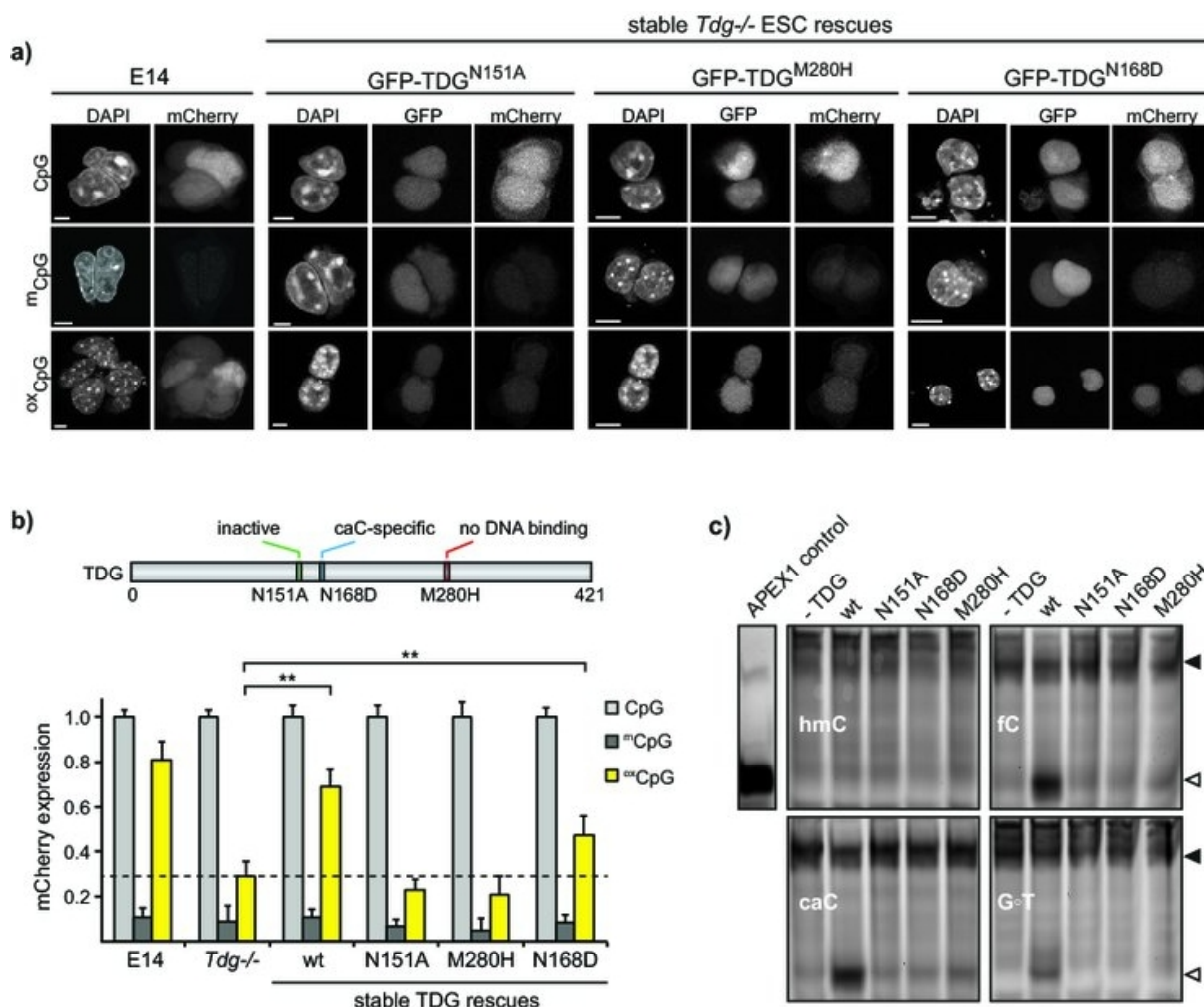
Additionally, we isolated genomic DNA from wt E14 and *Tdg*<sup>-/-</sup> ESCs as well as from the transient rescues with wt TDG and NEIL1, 2 and 3. Slot blot analyses were carried out for relative hmC, fC and caC quantifications. Genomic hmC was present at comparable levels in all tested cell lines and was not affected by *Tdg* knockout or NEIL overexpression. Since TDG is able to recognize and excise fC and caC, accumulation of these oxidized bases was observed in *Tdg*<sup>-/-</sup> ESCs, consistent with previous reports (30,52). Rescues of *Tdg*<sup>-/-</sup> cells with transiently expressed wt TDG, NEIL1, NEIL2 or NEIL3 resulted in decreased genomic fC, and for wt TDG and NEIL1, also in decreased caC levels (Figure 5c). These findings support the role of the glycosylases TDG, NEIL1, NEIL2 and NEIL3 in active DNA demethylation and subsequent reactivation of gene expression via excision of fC and caC followed by BER (Figure 6).

### DISCUSSION

In this study, we investigated the effects of the oxidized cytosine variants hmC, fC and caC on gene expression. By carrying out the enzymatic oxidation of a methylated



**Figure 3.** Oxidation of mCpG plasmid DNA leads to TDG-dependent gene reactivation. **(a)** *Tdg*<sup>-/-</sup> ESCs were transfected with pOct4-mCherry plasmids containing either unmodified, methylated or oxidized CpGs. Confocal images show a defect of oxCpG gene reactivation in *Tdg*<sup>-/-</sup> ESCs but not in *Mbd4*<sup>-/-</sup> ESCs. Transient rescue of *Tdg*<sup>-/-</sup> ESCs with GFP-TDG re-establishes oxCpG reporter gene expression. Cells were fixed with formaldehyde and counterstained with DAPI. Scale bar: 5  $\mu$ m. **(b)** High-throughput image acquisition and quantification of pOct4-mCherry expression shows that oxidation of mCpG sites in the pOct4-reporter results in reactivation of mCherry-expression in wt E14 ESCs and *Mbd4*<sup>-/-</sup> ESCs, but not in *Tdg*<sup>-/-</sup> ESCs. Expression of GFP-TDG rescues the phenotype (student's *t*-test,  $^{**}P < 0.025$ ,  $n = 200\ 000$ ; error bars indicate standard deviation). **(c)** Analytical digest with HpaII of differentially modified reporter plasmid DNA before and after transfection confirms substitution of oxCpG with CpG in wt E14 and *Mbd4*<sup>-/-</sup> ESCs.



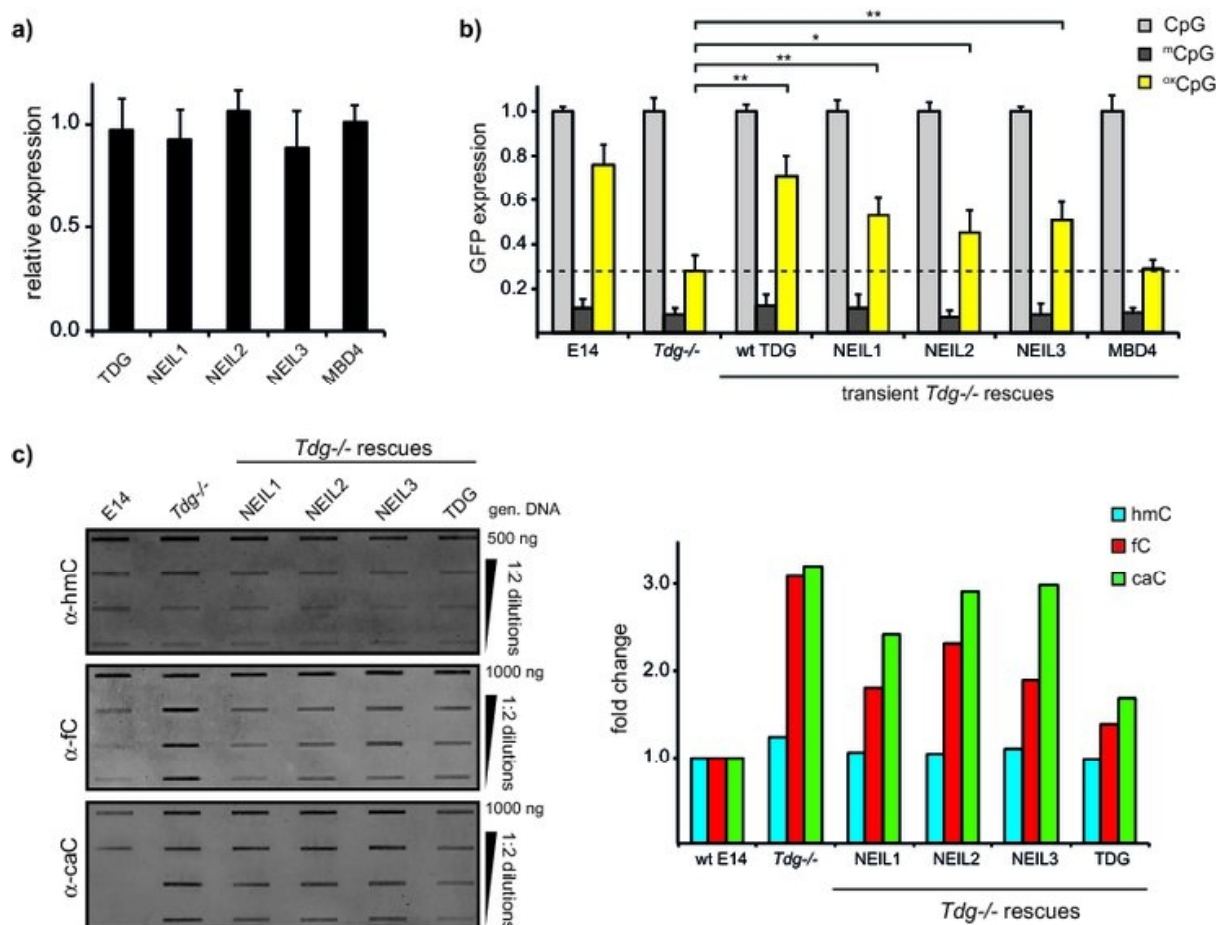
**Figure 4.** TDG activity is essential for gene reactivation. **(a)** Confocal images depicting expression levels of oxidized pOct4-mCherry expression in *Tdg*<sup>-/-</sup> ESCs stably rescued with GFP-TDG<sup>N151A</sup>, GFP-TDG<sup>M280H</sup> and GFP-TDG<sup>N168D</sup> in comparison to wt E14 ESCs. Scale bar: 5 μm. **(b)** High-throughput image acquisition and quantification of pOct4-mCherry expression in wt E14, *Tdg*<sup>-/-</sup> and *Tdg*<sup>-/-</sup> ESCs stably expressing wt, catalytically inactive, DNA binding deficient and caC-specific TDG mutants. Methylation of the pOct4-mCherry reporter leads to a 5-fold lower expression compared to unmodified plasmid. Oxidation of mCpG sites in the pOct4-reporter results in reactivation of mCherry-expression in wt ESCs but not in *Tdg*<sup>-/-</sup> ESCs. This re-increase was also obtained in *Tdg*<sup>-/-</sup> ESCs rescued with wt or caC-specific TDG, while the latter was not as efficient (student's *t*-test, \*\**P* < 0.025, *n* = 200 000; error bars indicate standard deviation). **(c)** TDG activity was monitored using an *in vitro* assay based on the ability of APEX1 to create a single strand break out of an abasic site, detectable as distinct band on the gel (empty triangle). Full length DNA is indicated by a solid triangle. The DNA substrates contain one defined modification site as indicated in the figure. Wt TDG is highly active on fC or caC and to a much lesser extent on a G-T mismatch.

pOct4-reporter construct *in vitro*, we separated the generation of modified cytosines from their further processing *in vivo*. This allowed us to directly investigate the cellular factors responsible for gene reactivation independent of endogenous TET activity and regulation. In wt ESCs, we observed strong reporter expression from oxidized but not from methylated plasmids, suggesting oxidation-dependent gene reactivation.

To investigate which pathway is responsible for the observed gene reactivation, we searched for potential TET interaction partners. So far, it has been shown that MBD3

colocalizes with TET1 regulating hmC-marked gene expression (53) and that TET1, TET2 and TET3 interact with OGT controlling protein stability, localization and histone modification (54–57). Also, several chromatin-binding factors, such as HDAC1, EZH2 and MeCP2, have been described to associate with TET1 (58). However, none of these factors is likely to be involved in the process of DNA demethylation. Therefore, we performed a mass spectrometry-based pull-down approach in which several BER factors co-precipitated with TET1. To confirm these results, we used Co-IP and a recently described F3H as-





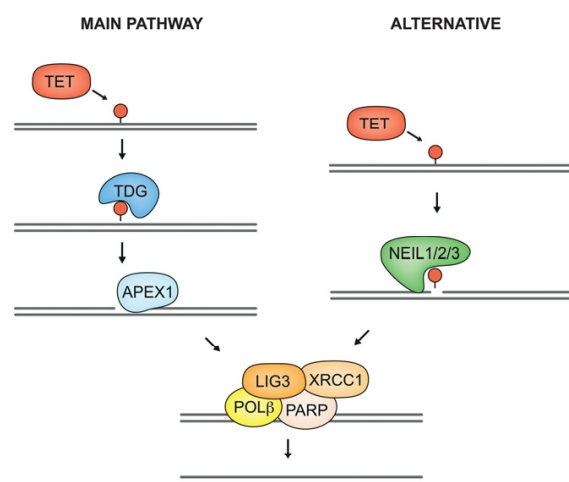
**Figure 5.** The NEIL glycosylase family can partially compensate for TDG. (a) Quantification of mCherry intensities with high-throughput imaging shows that transient *Tdg*<sup>-/-</sup> rescue ESCs express mCherry-tagged TDG, NEIL1/2/3 and MBD4 glycosylases at comparable levels ( $n = 100\,000$ ; error bars indicate standard deviation). (b) The ability of the NEIL family of glycosylases to substitute TDG *in vivo* was monitored by expression of differentially modified pOct4-GFP. With ectopic expression of mCherry-NEIL1, NEIL2 and NEIL3, reporter gene signal was significantly higher than in *Tdg*<sup>-/-</sup> cells, although not reaching the levels of the wt mCherry-TDG rescue. MBD4 overexpression could not compensate for loss of TDG (student's *t*-test, \* $P < 0.05$ , \*\* $P < 0.025$ ,  $n = 200\,000$ ; error bars indicate standard deviation). (c) Slot blot analysis of genomic DNA isolated from wt E14, *Tdg*<sup>-/-</sup> and the indicated rescues shows constant levels of hmC and accumulation of fC and caC in *Tdg*<sup>-/-</sup> cells. Overexpression of wt TDG or NEIL1, 2 or 3 leads to a decrease in genomic fC amounts but does not reach wt E14 levels. Expression of wt TDG or NEIL1 also reduces caC accumulation.

say to test interaction of all three TETs with different BER factors *in vitro* and *in vivo*. We were able to detect interactions of TET proteins with the DNA glycosylases TDG, MBD4 as well as NEIL1, NEIL2 and NEIL3, which excise damaged or oxidized DNA bases (18,25,32,59,60). Furthermore, interactions of all three TET proteins were observed with PARP1, which detects single-strand breaks and modifies repair factors by PolyADP-ribosylation (61). Finally, we showed TET interactions with LIG3 and XRCC1, which are recruited by PARP1 to the site of DNA damage and ligate the DNA strand after the insertion of cytosine (62–64). The observed interactions were largely independent of TET activity, indicating direct protein–protein interactions.

Our findings suggest that both TET-dependent oxidation of mC and subsequent excision of oxidized cytosines by the

BER machinery take place in one large protein complex in a spatially and temporally coordinated manner. This close association enables highly efficient replacement of oxidized cytosines. In accordance with these results, fC and caC, in contrast to hmC, are detected at very low genomic levels and are proposed to be immediately removed after generation (18,19).

Initial hypotheses proposed that hmC might be deaminated to hmU by AID/APOBEC deaminases prior to base excision by DNA glycosylases. Suggested candidates were TDG and MBD4, which both have been shown to recognize G•T and G•hmU mismatches (23,65) and have been identified as TET protein interactors in this study. TDG has also been described to be active on fC and caC (18,25). In contrast to wt ESCs, no gene reactivation on the oxidized



**Figure 6.** Two alternative pathways for TET-mediated active demethylation. TDG is the major glycosylase that removes fC or caC, generated by TET proteins. Alternatively, the NEIL glycosylases can excise oxidized cytosines, but less efficiently. Both pathways are completed by the BER machinery.

reporter plasmid was observed in *Tdg*<sup>-/-</sup> cells. Knockout of *Mbd4*, however, had no effect on reporter gene expression, although MBD4 interacts with TET1, TET2 and TET3. Apparently, this association does not contribute to gene activation and may be involved in a different regulatory mechanism.

Our data suggest that conversion of hmC to hmU and subsequent excision by MBD4 does not play a major role in ESCs and are in line with previous studies showing that AID is unable to operate on double-stranded DNA and no detectable deamination of hmC by AID/APOBEC *in vitro* or *in vivo* could be observed (34,66). However, we cannot rule out that AID/APOBEC is involved in the TDG-dependent demethylation pathway or that this pathway contributes to TET3-dependent active demethylation in zygotes (67). Restriction enzyme-based analysis of the oxidized plasmid DNA recovered from wt E14 ESCs provided evidence that conversion of oxidized cytosine to unmodified cytosine led to gene activation. Again, this effect was dependent on TDG but not on MBD4.

Additional *in vivo* experiments with different TDG mutants showed that TDG activity and not the recruitment of unknown factors through TDG is essential for the recovery of gene expression. The specificity of TDG toward fC and caC, but not hmC, was confirmed using a newly established assay based on the ability of APEX1 to recognize glycosylase-generated abasic sites and convert them into single-strand breaks. The results also revealed that TDG activity is much higher on fC or caC than on a G•T mismatch, arguing that oxidized cytosines are the major substrate for TDG and that deamination is not necessary for gene reactivation.

Besides TDG and MBD4, we also investigated the role of the NEIL glycosylase family in TET-mediated demethylation. NEIL1, NEIL2 and NEIL3 have been shown to excise

several lesions resulting from DNA oxidations, such as 5-hydroxyuracil or thymine glycol (59,68). NEIL glycosylases are bifunctional, i.e. are also able to convert abasic sites into single-strand breaks (69). *Neil3*<sup>-/-</sup> mice have been reported to be viable as well as fertile, and the expression of NEIL3 is elevated in hematopoietic tissues, suggesting a function in the immune system or hematopoiesis (70).

Since the NEIL glycosylases have been described as potential binders of oxidized cytosines (16), we tried to compensate for the loss of TDG in *Tdg*<sup>-/-</sup> cells with each of the three NEIL proteins. Indeed, we could detect a significant reactivation of reporter gene expression, although not reaching the levels of the rescue with wt TDG. We conclude that the NEIL glycosylases can also initiate BER after TET-mediated cytosine oxidation. This hypothesis was further confirmed by the fact that the accumulation of genomic fC and caC in *Tdg*<sup>-/-</sup> ESCs was far less prominent when NEIL1, 2 or 3 was overexpressed. These data clearly show that the NEIL proteins are not only capable of reactivating the oxidized reporter gene, but also of excising formylated and carboxylated cytosine in its chromatin context. Thus, the NEIL glycosylases may constitute an alternative pathway for active demethylation and reactivation of epigenetically silenced genes (Figure 6).

In summary, we show that the TET proteins interact with a set of factors involved in catalyzing the multiple steps of BER. Furthermore, we demonstrate that TDG is the main glycosylase in TET-mediated reactivation of the epigenetically silenced Oct4 promoter via the BER pathway. It would be of interest whether the activity of the TET-BER machinery differs on other promoter types, such as CpG island containing promoters. Our results also indicate that the NEIL family of glycosylases can functionally replace TDG. It remains to be elucidated to which extent the NEIL glycosylases contribute to TET-mediated demethylation and gene reactivation and how the usage of different glycosylases is regulated *in vivo*.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank Prof. P. Schär (University of Basel) for providing *Tdg*<sup>-/-</sup> ESC and N. Nagaraj at the core facility of the MPI for Biochemistry for mass spectrometry analysis. Furthermore, we want to thank E. Schmidtman for help with establishing the TDG/APEX1 activity assay, C. Trummer for F3H quantifications, D. Meilinger for FACS sorting and C. Mulholland for suggestions on the manuscript.

## FUNDING

Deutsche Forschungsgemeinschaft [Collaborative Research Centers SFB 646/B10 and SFB 1064/A17]. C.B. is a member of IMPRS-LS; H.L. is a member of the Nanosystems Initiative Munich (NIM).

*Conflict of interest.* None declared.

## REFERENCES

- Wutz, A. and Barlow, D.P. (1998) Imprinting of the mouse *Igf2r* gene depends on an intronic CpG island. *Mol. Cell. Endocrinol.*, **140**, 9–14.
- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
- Bestor, T., Laudano, A., Mattaliano, R. and Ingram, V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.*, **203**, 971–983.
- Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
- Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B. and Bestor, T.H. (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science*, **294**, 2536–2539.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
- Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
- Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. and Leonhardt, H. (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.*, **38**, e181.
- Iyer, L.M., Tahiliani, M., Rao, A. and Aravind, L. (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, **8**, 1698–1710.
- Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S. et al. (2013) NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature*, **495**, 370–374.
- Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunliffe, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G. et al. (2011) Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, **8**, 200–213.
- Piccolo, F.M., Bagci, H., Brown, K.E., Landeira, D., Soza-Ried, J., Feytout, A., Mooijman, D., Hajkova, P., Leitch, H.G., Tada, T. et al. (2013) Different roles for Tet1 and Tet2 proteins in reprogramming-mediated erasure of imprints induced by EGC fusion. *Mol. Cell*, **49**, 1023–1033.
- Konstandin, N., Bultmann, S., Szwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P.M., Schneider, S. et al. (2011) Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia*, **25**, 1649–1652.
- Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R. et al. (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, **468**, 839–843.
- Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F. et al. (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146–1159.
- Kellinger, M.W., Song, C.X., Chong, J., Lu, X.Y., He, C. and Wang, D. (2012) 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.*, **19**, 831–833.
- Maiti, A. and Drohat, A.C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.*, **286**, 35334–35338.
- Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C.A., Hagemeyer, C. and Carell, T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie*, **50**, 7008–7012.
- Inoue, A. and Zhang, Y. (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science*, **334**, 194.
- Iqbal, K., Jin, S.G., Pfeifer, G.P. and Szabo, P.E. (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 3642–3647.
- Cardoso, M.C. and Leonhardt, H. (1999) DNA methyltransferase is actively retained in the cytoplasm during early development. *J. Cell Biol.*, **147**, 25–32.
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L. and Song, H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, **145**, 423–434.
- Wu, S.C. and Zhang, Y. (2010) Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.*, **11**, 607–620.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L. et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303–1307.
- Campalans, A., Kortulewski, T., Amouroux, R., Menoni, H., Vermeulen, W. and Radicella, J.P. (2013) Distinct spatiotemporal patterns and PARP dependence of XRCC1 recruitment to single-strand break and base excision repair. *Nucleic Acids Res.*, **41**, 3115–3129.
- Dianova, I.I., Sleeth, K.M., Allinson, S.L., Parsons, J.L., Breslin, C., Caldecott, K.W. and Dianov, G.L. (2004) XRCC1-DNA polymerase beta interaction is required for efficient base excision repair. *Nucleic Acids Res.*, **32**, 2550–2555.
- Mortusewicz, O., Rothbauer, U., Cardoso, M.C. and Leonhardt, H. (2006) Differential recruitment of DNA Ligase I and III to DNA repair sites. *Nucleic Acids Res.*, **34**, 3523–3532.
- Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F. et al. (2011) Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*, **470**, 419–423.
- Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.L., Zhang, K. and Zhang, Y. (2013) Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*, **153**, 692–706.
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D. et al. (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, **146**, 67–79.
- Hashimoto, H., Zhang, X. and Cheng, X. (2012) Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res.*, **40**, 8276–8284.
- Kemmerich, K., Dingler, F.A., Rada, C. and Neuberger, M.S. (2012) Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice. *Nucleic Acids Res.*, **40**, 6016–6025.
- Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y. and Kohli, R.M. (2012) AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.*, **8**, 751–758.
- Schiesser, S., Hackner, B., Pfaffeneder, T., Muller, M., Hagemeyer, C., Truss, M. and Carell, T. (2012) Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angewandte Chemie*, **51**, 6516–6520.
- Tsukamoto, T., Hashiguchi, N., Janicki, S.M., Tumber, T., Belmont, A.S. and Spector, D.L. (2000) Visualization of gene activity in living cells. *Nat. Cell Biol.*, **2**, 871–878.
- Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R. and Bird, A. (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science*, **297**, 403–405.
- Ying, Q.L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008) The ground state of embryonic stem cell self-renewal. *Nature*, **453**, 519–523.
- Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C. and Zhang, Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion,

- ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
40. Szwagierczak, A., Brachmann, A., Schmidt, C.S., Bultmann, S., Leonhardt, H. and Spada, F. (2011) Characterization of PvuRtsII endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **39**, 5149–5156.
  41. Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C. and Leonhardt, H. (2008) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol. Cell. Proteomics*, **7**, 282–289.
  42. Rottach, A., Kremmer, E., Nowak, D., Leonhardt, H. and Cardoso, M.C. (2008) Generation and characterization of a rat monoclonal antibody specific for multiple red fluorescent proteins. *Hybridoma*, **27**, 337–343.
  43. Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
  44. Herce, H.D., Deng, W., Helma, J., Leonhardt, H. and Cardoso, M.C. (2013) Visualization and targeted disruption of protein interactions in living cells. *Nat. Commun.*, **4**, 2660.
  45. Zolghadr, K., Rothbauer, U. and Leonhardt, H. (2012) The fluorescent two-hybrid (F2H) assay for direct analysis of protein-protein interactions in living cells. *Methods Mol. Biol.*, **812**, 275–282.
  46. Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X. and Cheng, X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.*, **40**, 4841–4849.
  47. Hardeland, U., Bentele, M., Jiricny, J. and Schar, P. (2000) Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J. Biol. Chem.*, **275**, 33449–33456.
  48. Hashimoto, H., Zhang, X. and Cheng, X. (2013) Selective excision of 5-carboxylcytosine by a thymine DNA glycosylase mutant. *J. Mol. Biol.*, **425**, 971–976.
  49. Hashimoto, H., Hong, S., Bhagwat, A.S., Zhang, X. and Cheng, X. (2012) Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res.*, **40**, 10203–10214.
  50. Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J.J., Wiebauer, K. and Jiricny, J. (1996) Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem.*, **271**, 12767–12774.
  51. Wiebauer, K. and Jiricny, J. (1990) Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 5842–5845.
  52. Raiber, E.A., Beraldi, D., Ficiz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W. and Balasubramanian, S. (2012) Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.*, **13**, R69.
  53. Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J. and Fazio, T.G. (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*, **147**, 1498–1510.
  54. Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Mendez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E. et al. (2013) TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.*, **32**, 645–655.
  55. Chen, Q., Chen, Y., Bian, C., Fujiki, R. and Yu, X. (2013) TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature*, **493**, 561–564.
  56. Shi, F.T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M.A., Wan, M. and Songyang, Z. (2013) Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *J. Biol. Chem.*, **288**, 20776–20784.
  57. Zhang, Q., Liu, X., Gao, W., Li, P., Hou, J., Li, J. and Wong, J. (2014) Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked beta-N-acetylglucosamine transferase (OGT). *J. Biol. Chem.*, **289**, 5986–5996.
  58. Cartron, P.F., Nadaradjane, A., Lepape, F., Lallier, L., Gardie, B. and Vallette, F.M. (2013) Identification of TET1 Partners That Control Its DNA-Demethylating Function. *Genes Cancer*, **4**, 235–241.
  59. Dou, H., Mitra, S. and Hazra, T.K. (2003) Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J. Biol. Chem.*, **278**, 49679–49684.
  60. Hazra, T.K., Kow, Y.W., Hatahet, Z., Imhoff, B., Boldogh, I., Mookapathi, S.K., Mitra, S. and Izumi, T. (2002) Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J. Biol. Chem.*, **277**, 30417–30420.
  61. Gibson, B.A. and Kraus, W.L. (2012) New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nat. Rev. Mol. Cell Biol.*, **13**, 411–424.
  62. Kubota, Y., Nash, R.A., Klungland, A., Schar, P., Barnes, D.E. and Lindahl, T. (1996) Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. *EMBO J.*, **15**, 6662–6670.
  63. Rice, P.A. (1999) Holding damaged DNA together. *Nat. Struct. Biol.*, **6**, 805–806.
  64. Mortusewicz, O., Ame, J.C., Schreiber, V. and Leonhardt, H. (2007) Feedback-regulated poly(ADP-ribosylation) by PARP-1 is required for rapid response to DNA damage in living cells. *Nucleic Acids Res.*, **35**, 7665–7675.
  65. Niehrs, C. and Schafer, A. (2012) Active DNA demethylation by Gadd45 and DNA repair. *Trends Cell Biol.*, **22**, 220–227.
  66. Branstetter, R., Pham, P., Scharff, M.D. and Goodman, M.F. (2003) Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 4102–4107.
  67. Santos, F., Peat, J., Burgess, H., Rada, C., Reik, W. and Dean, W. (2013) Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenet. Chromat.*, **6**, 39.
  68. Hazra, T.K., Izumi, T., Boldogh, I., Imhoff, B., Kow, Y.W., Jaruga, P., Dizdaroğlu, M. and Mitra, S. (2002) Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 3523–3528.
  69. Jacobs, A.L. and Schar, P. (2012) DNA glycosylases: in DNA repair and beyond. *Chromosoma*, **121**, 1–20.
  70. Torisu, K., Tsuchimoto, D., Ohnishi, Y. and Nakabeppu, Y. (2005) Hematopoietic tissue-specific expression of mouse Neil3 for endonuclease VIII-like protein. *J. Biochem.*, **138**, 763–772.

## **SUPPLEMENTARY FIGURES**

TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation

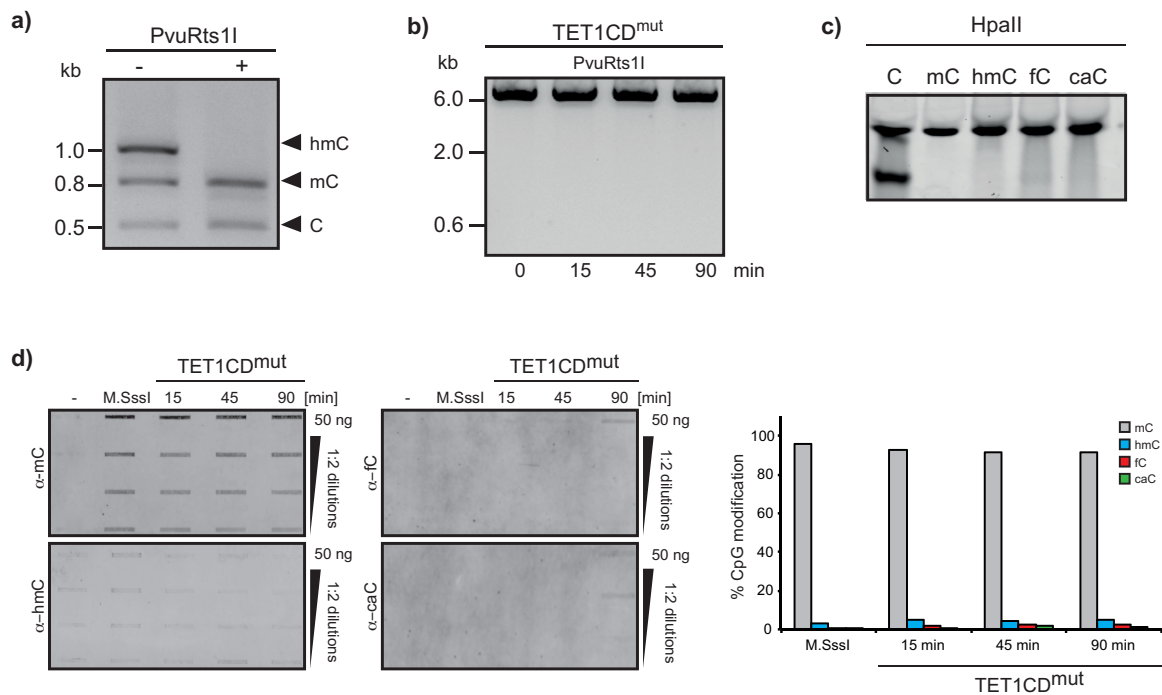
Udo Müller, Christina Bauer, Michael Siegl, Andrea Rottach, Heinrich Leonhardt

Department of Biology II, Ludwig Maximilians University Munich and Center for Integrated Protein Science Munich (CIPSM), 82152 Planegg-Martinsried, Germany



## 2. Results

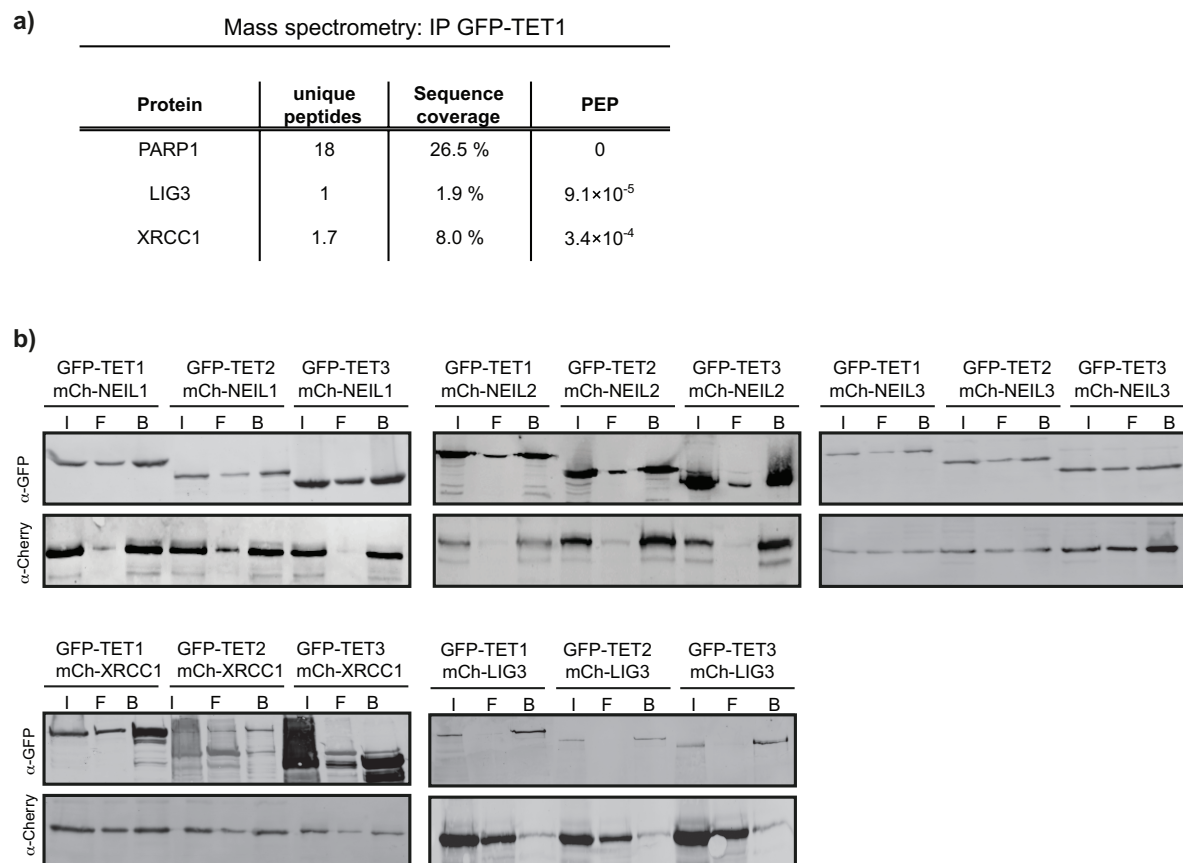
Supplementary Figure S1



### Supplementary Figure S1:

- a)** PvuRts1I specifically cuts hmC-containing DNA. PCR fragments were generated using dATP, dTTP and dGTP with either dCTP, dmCTP or dhmCTP.
- b)** Incubation of methylated pOct4-GFP with TET1CD<sup>mut</sup> does not result in any PvuRts1I digestion showing that no DNA oxidation takes place.
- c)** HpaII specifically cuts unmodified cytosine in a CCGG context. Substrate: 42 bp oligo, containing one defined modification site.
- d)** Cytosine modifications of untreated, M.SssI-methylated and TET1CD<sup>mut</sup>-incubated pOct4-mCherry plasmid DNA are detected by slot blot. A 2fold serial dilution of the plasmid DNA was loaded and detected using antibodies against mC, hmC, fC and caC. The mC levels remain high for all samples, except for untreated DNA. Almost no detectable increase in hmC, fC and caC can be observed.

## Supplementary Figure S2

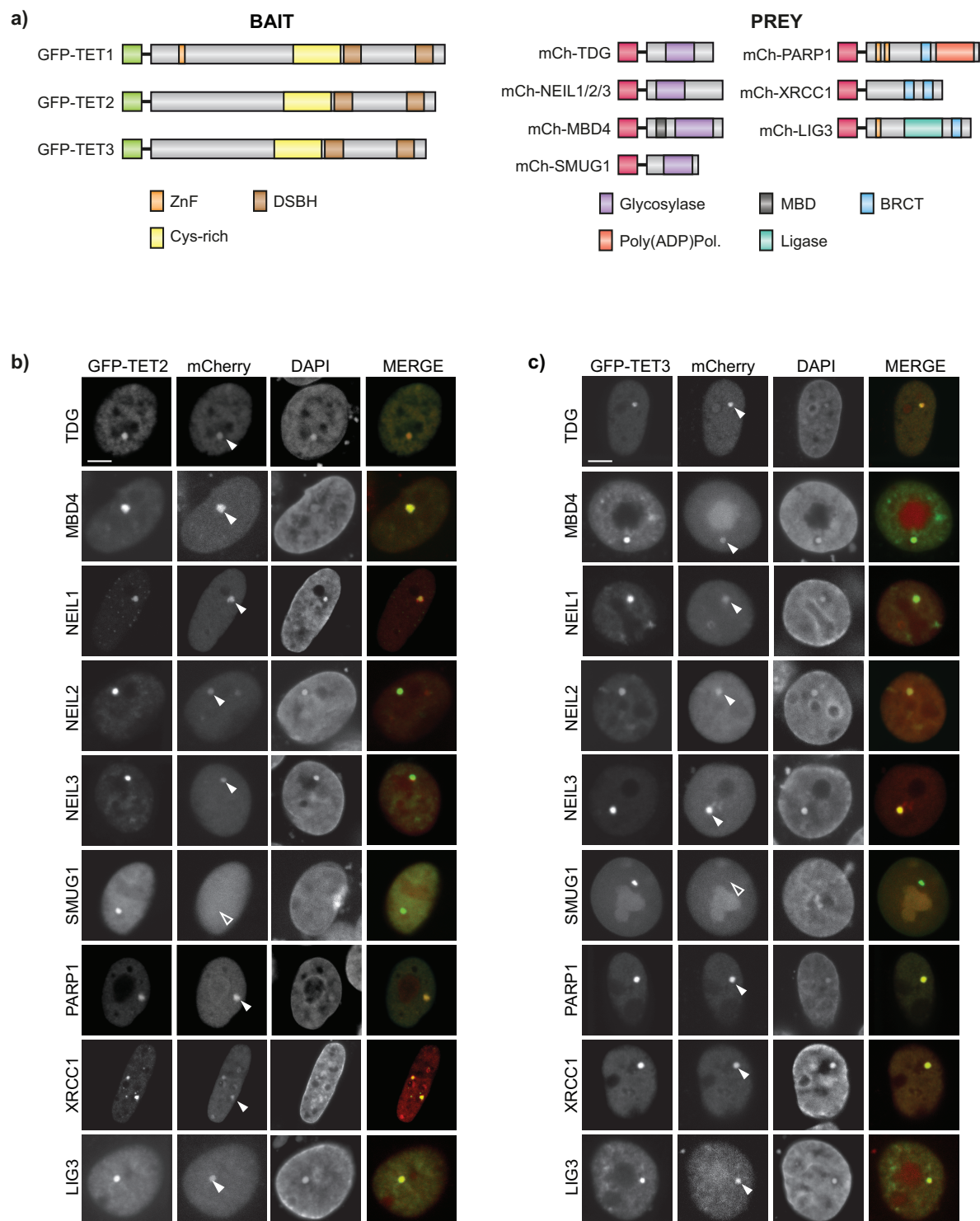
**Supplementary Figure S2:**

**a)** PARP1, XRCC1 and LIG3 have been identified as interactors of TET1 by mass spectrometry. Numbers represent the average of three biological replicates. PEP: Posterior error probability.

**b)** Co-immunoprecipitation with subsequent SDS-PAGE and Western Blotting shows interaction of all three TET proteins with the NEIL1, NEIL2 and NEIL3 glycosylases as well as XRCC1 and LIG3. GFP-tagged TET proteins and the respective mCherry-fusions were expressed in HEK293T cells and immunoprecipitated using the GFP-Trap (I: Input, F: Flowthrough, B: Bound).

## 2. Results

Supplementary Figure S3



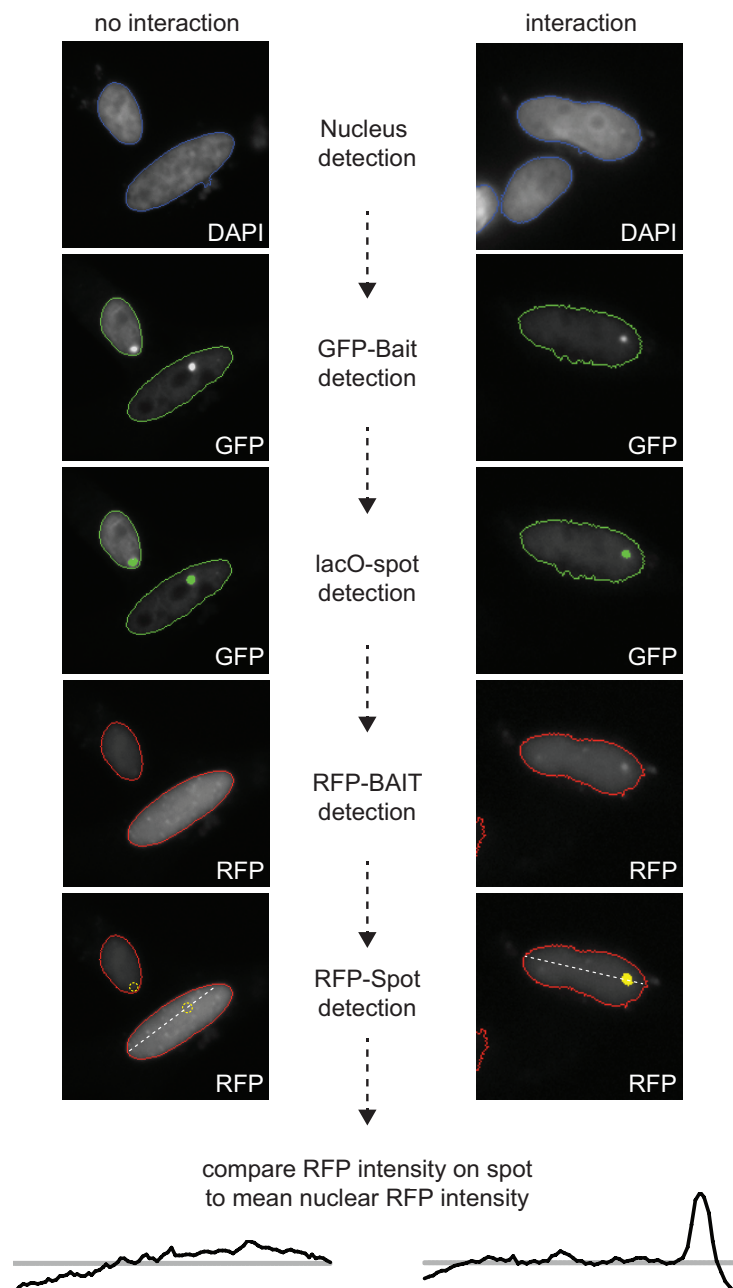
**Supplementary Figure S3:**

**a)** Schematic representation of the GFP-BAIT and mCherry-PREY constructs used to analyze TET-BER interactions in F3H and Co-IP.

**b, c)** F3H assay shows *in vivo* interactions of TET2 and TET3 with various glycosylases and BER factors. Positive and negative interactions are marked with a solid or empty triangle, respectively. Scale bar: 5  $\mu$ m



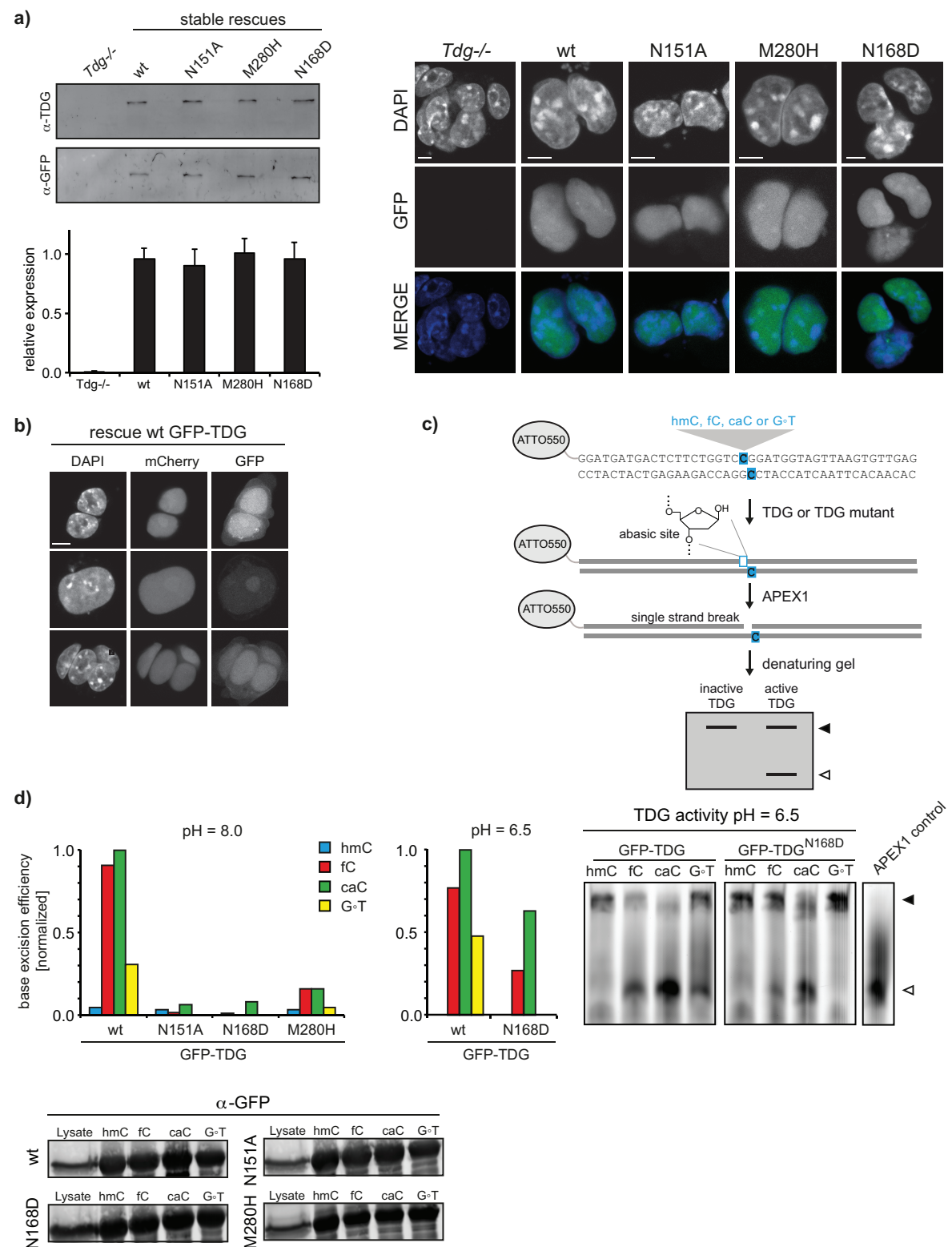
Supplementary Figure S4

**Supplementary Figure S4:**

Schematic depiction of the workflow used for automated image acquisition and quantification of the F3H assay.

## 2. Results

Supplementary Figure S5



**Supplementary Figure S5:**

**a)** Stable *Tdg*<sup>-/-</sup> rescue ESCs express GFP-TDG and mutants at comparable levels as shown by Western blot analysis (upper left), high-throughput imaging (lower left) and confocal microscopy (right) (n=100,000; error bars indicate standard deviation). Scale bar: 5 μm

- b)** Rescue of *Tdg*<sup>-/-</sup> ESCs with wt GFP-TDG re-establishes the capability to express the reporter gene from the <sup>ox</sup>CpG plasmid. Scale bar: 5  $\mu$ m
- c)** Schematic representation of the APEX1-based TDG activity assay: if TDG is active the formed abasic site is converted into a single strand break by APEX1, resulting in two DNA fragments on a denaturing gel.
- d)** Quantification of the *in vitro* base-excision activity of purified TDG on defined DNA substrates (left). Equal amounts of GFP-TDG and the mutants were incubated with a 42 bp DNA oligonucleotide harboring a central modified CpG site. At pH=6.5, the caC-specific TDG mutant (N168D) displays a preference for caC and, to a lesser extent, towards fC (right). To confirm equivalent protein levels, protein signal was immunodetected by Western blot using a mouse-anti-GFP antibody (Roche) (bottom).

## 2. Results

### **2.3 Phosphorylation of TET proteins is regulated via *O*-GlcNAcylation by the *O*-linked *N*-Acetylglucosamine transferase (OGT)**

published January 7, 2015, in *The Journal of Biological Chemistry*

# Phosphorylation of TET Proteins Is Regulated via O-GlcNAcylation by the O-Linked N-Acetylglucosamine Transferase (OGT)<sup>\*[S]</sup>

Received for publication, August 21, 2014, and in revised form, December 19, 2014. Published, JBC Papers in Press, January 7, 2015, DOI 10.1074/jbc.M114.605881

Christina Bauer<sup>†1</sup>, Klaus Göbel<sup>‡</sup>, Nagarjuna Nagaraj<sup>§</sup>, Christian Colantuoni<sup>‡</sup>, Mengxi Wang<sup>‡</sup>, Udo Müller<sup>‡</sup>, Elisabeth Kremmer<sup>¶</sup>, Andrea Rottach<sup>‡2</sup>, and Heinrich Leonhardt<sup>‡||3</sup>

From the <sup>†</sup>Biocenter, Ludwig-Maximilians University Munich, 82152 Planegg-Martinsried, the <sup>§</sup>Max Planck Institute of Biochemistry, D-82152 Martinsried, the <sup>¶</sup>Institute for Molecular Immunology, Helmholtz Center Munich, 81377 München-Großhadern, and the <sup>||</sup>Center for Integrated Protein Science Munich (CIPSM), 81377 München, Germany

**Background:** TET proteins oxidize 5-methylcytosine and contribute to active DNA demethylation.

**Results:** O-Linked GlcNAc transferase modifies TET proteins with GlcNAc and thereby reduces TET phosphorylation.

**Conclusion:** TET proteins are subjected to a dynamic interplay of post-translational modifications at low-complexity regions.

**Significance:** This first map of TET phosphorylation and O-GlcNAcylation sites at amino acid resolution provides a valuable resource for future studies of TET regulation.

TET proteins oxidize 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine and thus provide a possible means for active DNA demethylation in mammals. Although their catalytic mechanism is well characterized and the catalytic dioxygenase domain is highly conserved, the function of the regulatory regions (the N terminus and the low-complexity insert between the two parts of the dioxygenase domains) is only poorly understood. Here, we demonstrate that TET proteins are subject to a variety of post-translational modifications that mostly occur at these regulatory regions. We mapped TET modification sites at amino acid resolution and show for the first time that TET1, TET2, and TET3 are highly phosphorylated. The O-linked GlcNAc transferase, which we identified as a strong interactor with all three TET proteins, catalyzes the addition of a GlcNAc group to serine and threonine residues of TET proteins and thereby decreases both the number of phosphorylation sites and site occupancy. Interestingly, the different TET proteins display unique post-translational modification patterns, and some modifications occur in distinct combinations. In summary, our results provide a novel potential mechanism for TET protein regulation based on a dynamic interplay of phosphorylation and O-GlcNAcylation at the N terminus and the low-complexity insert region. Our data suggest strong cross-talk between the modification sites that could allow rapid adaption of TET protein localization, activity, or targeting due to changing environmental conditions as well as in response to external stimuli.

A major epigenetic mechanism of gene regulation in higher eukaryotes is methylation of DNA at C5 of cytosines (1, 2). Recently, the family of TET (ten-eleven translocation) proteins has been shown to successively oxidize 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine (3–6), providing novel insights into the dynamics of DNA modifications. TET proteins are also active on genomic thymine residues, leading to the generation of 5-hydroxyuracil (7). In Gnathostomata, there are three TET proteins, TET1, TET2, and TET3 (8), which show distinct expression patterns and functions in different tissues or during development (9–13). TET1 and TET2 are highly expressed in mouse embryonic stem cells (mESCs)<sup>4</sup> and are associated with oxidation of transcription start sites and gene bodies, respectively (14). TET3 is up-regulated in the oocyte and oxidizes the silenced paternal pronuclear DNA (10, 15). High levels of TET proteins and genomic 5-hydroxymethylcytosine are described for neuronal tissues (11, 16–18). In several patients with myeloid malignancies, mutations of *TET2* correlate with decreased 5-hydroxymethylcytosine levels and altered gene expression patterns (19–22).

The activity of TET proteins directly depends on two cofactors: Fe(II) and 2-oxoglutarate (3, 8). Interestingly, gain-of-function mutations of the enzymes responsible for 2-oxoglutarate synthesis, IDH1 and IDH2, have been associated with tumorigenesis, in particular glioblastomata and acute myeloid leukemia (20, 23, 24). These mutations lead to the synthesis of 2-hydroxyglutarate, a potent inhibitor of 2-oxoglutarate-dependent dioxygenases such as TET proteins (24, 25). Because IDH1 and IDH2 are enzymes of the Krebs cycle, these findings represent a direct link of TET protein activity to metabolism, especially because low 5-hydroxymethylcytosine levels are found in acute myeloid leukemia patients not only with *TET2* loss-of-function mutations but also with *IDH2* gain-of-function mutations (20). Besides 2-hydroxyglutarate, ascorbate has also been shown to influence cytosine oxidation by TET pro-

<sup>\*</sup> This work was supported in part by Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Center Grants SFB 646/B10 and SFB 1064/A17.

✂ Author's Choice—Final version full access.

[S] This article contains supplemental Data S1, S2, and S3 and supplemental Table S4.

<sup>1</sup> Supported by the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS).

<sup>2</sup> To whom correspondence may be addressed. E-mail: a.rottach@lmu.de.

<sup>3</sup> Member of the Nanosystems Initiative Munich (NIM). To whom correspondence may be addressed: Dept. of Biology II, Ludwig-Maximilians University Munich, Großhadernerstr. 2, 81925 Planegg-Martinsried, Germany. Tel.: 49-89-2180-74229; Fax: 49-89-2180-74236; E-mail: h.leonhardt@lmu.de.

<sup>4</sup> The abbreviations used are: mESC, mouse embryonic stem cell; OGT, O-linked GlcNAc transferase; PTM, post-translational modification.

### Phosphorylation and O-GlcNAcylation of TET Proteins

teins (26–28). In summary, TET protein activity appears to be modulated by several small molecules, either inhibitory such as 2-hydroxyglutarate or stimulating such as ascorbate.

TET proteins are influenced not only by certain metabolites but also by interacting proteins. TET1 forms complexes with heterochromatin-associated proteins such as HDAC1, HDAC2, SIN3A, and EZH2 (29). All three TET proteins interact with a variety of factors of the base-excision repair pathway, including PARP1, LIG3, and XRCC1, and also with several DNA glycosylases, including thymine-DNA glycosylase, NEIL1, and MDB4 (30). Another known interactor of TET proteins is the glycosyltransferase OGT (31–36), which represents an additional interesting connection with metabolism. OGT catalyzes the addition of a GlcNAc group to serine or threonine residues of target proteins (37). Its activity is dependent on the availability of a variety of metabolic molecules such as glucose, ATP, glutamine, and acetyl-CoA (38). The association of OGT with TET proteins has been reported to influence histone modifications and gene expression (31, 36), TET1 protein stability (33) and activity (34), and TET3 subcellular localization (35).

TET protein activity is widely studied in the context of development, tumorigenesis, and metabolic conditions. However, only very little is known about the structure and function of the non-catalytic domains of TET proteins. In this study, we show that TET proteins are subject to a large number of post-translational modifications (PTMs), predominantly occurring at the two low-complexity regions, which display only little sequence conservation: the N terminus and the insert region that separates the two parts of the catalytic dioxygenase domain and is predicted to be unstructured (8). We demonstrate that TET proteins are phosphorylated and that this phosphorylation can be suppressed via O-GlcNAcylation by the glycosyltransferase OGT. Detailed mapping of modification sites to the protein sequence shows that mostly the N terminus and insert region of TET proteins are subjected to PTMs and that their regulation depends on a dynamic interplay of different PTMs.

### EXPERIMENTAL PROCEDURES

**Antibody Generation**—A His-tagged protein fragment from the insert region of each TET protein (see Fig. 1a) was expressed in *Escherichia coli* BL21(DE3) cells (Novagen, Darmstadt, Germany) and purified with the TALON Superflow metal affinity resin system (Clontech, Saint Germain, France) under native conditions as described previously (39). Amino acids 1682–1914 for TET1, amino acids 1332–1779 for TET2, and amino acids 976–1521 for TET3 were used as antigens. Approximately 100  $\mu$ g of each antigen was injected both intraperitoneally and subcutaneously into Lou/C rats using CPG2006 (TIB MOLBIOL, Berlin, Germany) as adjuvant. After 8 weeks, the immune response was boosted intraperitoneally and subcutaneously 3 days before fusion. Fusion of the myeloma cell line P3X63-Ag8.653 with rat immune spleen cells was performed using PEG 1500 (Roche Diagnostics Deutschland GmbH, Mannheim, Germany). After fusion, the cells were cultured in 96-well plates using RPMI 1640 medium with 20% fetal calf serum, penicillin/streptomycin, pyruvate, and nonessential amino acids (PAA, Linz, Austria) supplemented with aminopterin (Sigma). Hybridoma supernatants were tested in a

solid-phase immunoassay. Microtiter plates were coated overnight with His-tagged TET antigens at a concentration of 3–5  $\mu$ g/ml in 0.1 M sodium carbonate buffer (pH 9.6). After blocking with nonfat milk (Frema Reform, granoVita, Heimertingen, Germany), hybridoma supernatants were added. Bound rat monoclonal antibodies were detected with a mixture of biotinylated mouse monoclonal antibodies against rat IgG heavy chains, avoiding anti-IgM monoclonal antibodies (anti-IgG1, anti-IgG2a, and anti-IgG2b (American Type Culture Collection, Manassas, VA) and anti-IgG2c (Ascension GmbH, Munich, Germany)). The biotinylated monoclonal antibodies were visualized with peroxidase-labeled avidin (Alexis, San Diego, CA) and *o*-phenylenediamine as chromogen in the peroxidase reaction. Anti-TET1 (clones 5D6, 5D8, 2H9, and 4H7; rat IgG2a), anti-TET2 (clone 9F7; rat IgG2a), and anti-TET3 (clones 11B6 and 23B9; rat IgG2a) antibodies were stably subcloned and further characterized (see Fig. 1b).

**mESC Culture, Co-immunoprecipitation, and MS/MS Analysis**—mESCs (J1) were cultured as described previously (9). Endogenous TET1 and TET2 proteins were pulled out via monoclonal antibodies (clones 5D6, 5D8, and 9F7) coupled to protein G-Sepharose beads as described (39). After co-immunoprecipitation, protein samples were digested on beads with trypsin according to standard protocols. Peptide mixtures were analyzed by electrospray MS/MS spectrometry. Experiments were performed with an LTQ Orbitrap XL mass spectrometer (Thermo Scientific, Waltham, MA). Spectra were analyzed with Mascot software (Matrix Science, Boston, MA).

**Expression Constructs**—Expression constructs for GFP-TET1, GFP-TET2, GFP-TET3, GFP, and Cherry were described previously (40–42). To generate the Cherry-OGT construct, the coding sequence was amplified using cDNA from E14 mESCs as template and subcloned into the pCAG-Cherry-1B vector. Expression constructs for Cherry-OGT(H508A) (hereafter referred to as OGT<sup>mut</sup>) were generated by overlap extension PCR. All constructs were verified by DNA sequencing (Eurofins Genomics, Ebersberg, Germany).

**HEK293T Culture, Co-immunoprecipitation, and Western Blot Analysis**—Co-immunoprecipitation followed by Western blotting with GFP- and Cherry-tagged proteins expressed in HEK293T cells was performed as described previously (30). O-GlcNAc was detected with a mouse monoclonal antibody (RL2, Abcam, Cambridge, United Kingdom) and an Alexa 647N-conjugated secondary antibody (Sigma).

**Sample Preparation for Mass Spectrometric Analysis**—All experiments were performed in biological triplicates. GFP-tagged TET proteins and/or Cherry-tagged OGT and OGT<sup>mut</sup> were expressed in HEK293T cells. Cell lysis with radioimmune precipitation assay buffer and immunoprecipitation with the GFP-Trap (ChromoTek GmbH, Martinsried, Germany) were performed as described previously (30). After immunoprecipitation, samples on beads were rinsed two times with wash buffer (20 mM Tris-HCl (pH 7.5), 300 mM NaCl, and 0.5 mM EDTA) and two times with immunoprecipitation buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 0.5 mM EDTA).

100  $\mu$ l of denaturation buffer (6 M guanidine hydrochloride, 10 mM tris(2-carboxyethyl)phosphine, and 40 mM chloroacetamide in 100 mM Tris (pH 8.5)) was added to the beads and heated



## Phosphorylation and O-GlcNAcylation of TET Proteins

at 70 °C for 5 min. The samples were then subjected to sonication in a Diagenode Bioruptor Plus system (UCD-300-TO) at maximum power settings for 10 cycles consisting of a 30-s pulse and 30-s pause. Following sonication, the samples were diluted 1:10 with digestion buffer (25 mM Tris (pH 8.5) containing 10% acetonitrile) and mixed by vortexing prior to enzyme digestion. Each sample was digested with 1 µg of endoproteinase Lys-C (Wako Chemicals, Neuss, Germany) for 4 h with subsequent digestion using 1 µg of trypsin (Promega, Madison, WI) under gentle rotation at 37 °C. After digestion, the samples were placed in a SpeedVac concentrator for 10 min to remove acetonitrile from the sample before StageTip purification using SDB-XC material (43). Peptides were then eluted from the StageTip and placed in the SpeedVac concentrator to reduce the sample volume to ~6 µl, and 5 µl of the sample was injected onto the column for MS/MS analysis.

**LC-MS/MS and Data Analysis**—Samples were loaded onto a column (15-cm length and 75-µm inner diameter; New Objective, Woburn, MA) packed with 3-µm ReproSil C<sub>18</sub> beads (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) using an EASY-nLC autosampler (Thermo Scientific) coupled via a nanoelectrospray source to the LTQ Orbitrap XL mass spectrometer. Each sample was analyzed using a 2-h reversed-phase gradient and a top 5 method for data-dependent acquisition. Full scans were acquired in the Orbitrap mass spectrometer after accumulating up to  $1 \times 10^6$  charges, and MS/MS with the five most abundant precursors was performed using low-energy ion-trap collision-induced dissociation. MS/MS spectra were recorded using the ion trap by radial ejection.

All raw files were analyzed using the MaxQuant computational proteomics platform (version 1.4.1.6) (44). Peak lists were searched with an initial mass deviation of 7 ppm and fragment ion deviation of 0.5 Thomson. Carbamidomethylation was used as a fixed modification. Oxidation of methionine; phosphorylation of serine, threonine, and tyrosine; O-GlcNAcylation of serine and threonine; ubiquitination (diglycine motif) of lysine; and acetylation of the protein N terminus were used as variable modifications. All unmodified and oxidized methionine- and N-acetylation-containing peptides were used for protein quantification. The MaxQuant software quantifies the different versions of modified peptides in a label-free fashion. Briefly, the occupancy reflects the extracted signal differences between modified and unmodified peptides and also includes the protein ratios between samples. The different forms of modified peptides, *e.g.* peptides with single, double, and triple O-GlcNAc sites, are individually quantified and listed separately in the output (supplemental Table S4). Details on label-free quantification of modification sites are provided elsewhere (45).

MaxQuant output data were further analyzed with Perseus software (version 1.5.0.15) (44). Only modifications that were detected in at least two of three biological replicates in at least one experimental setup were included in the analysis. PTMs that were detected in non-unique peptides were also excluded. Significance was tested using a Student's two-tailed paired *t* test.

## RESULTS

**Characterization of Anti-TET Antibodies**—The three TET proteins share a common domain architecture: the C-terminal catalytic dioxygenase domain is split into two parts separated by a low-complexity insert region and is preceded by an extension enriched in cysteines (8). All three TET proteins have a large N-terminal region that is mostly uncharacterized so far, except for a CXXC-type zinc finger at the N terminus of TET1 and TET3 (8, 40, 47). Murine TET3 exists in two isoforms: one with the zinc finger and one without (41). The cysteine-rich region and the split dioxygenase domain are conserved among the three murine TET proteins, whereas the N terminus and insert region display only little sequence similarity (Fig. 1*a*). The three-dimensional structure of mammalian TET proteins remains unresolved, with the exception of the cysteine-rich and dioxygenase domains of TET2 (48), leaving the structure and function of the N terminus and low-complexity insert unknown.

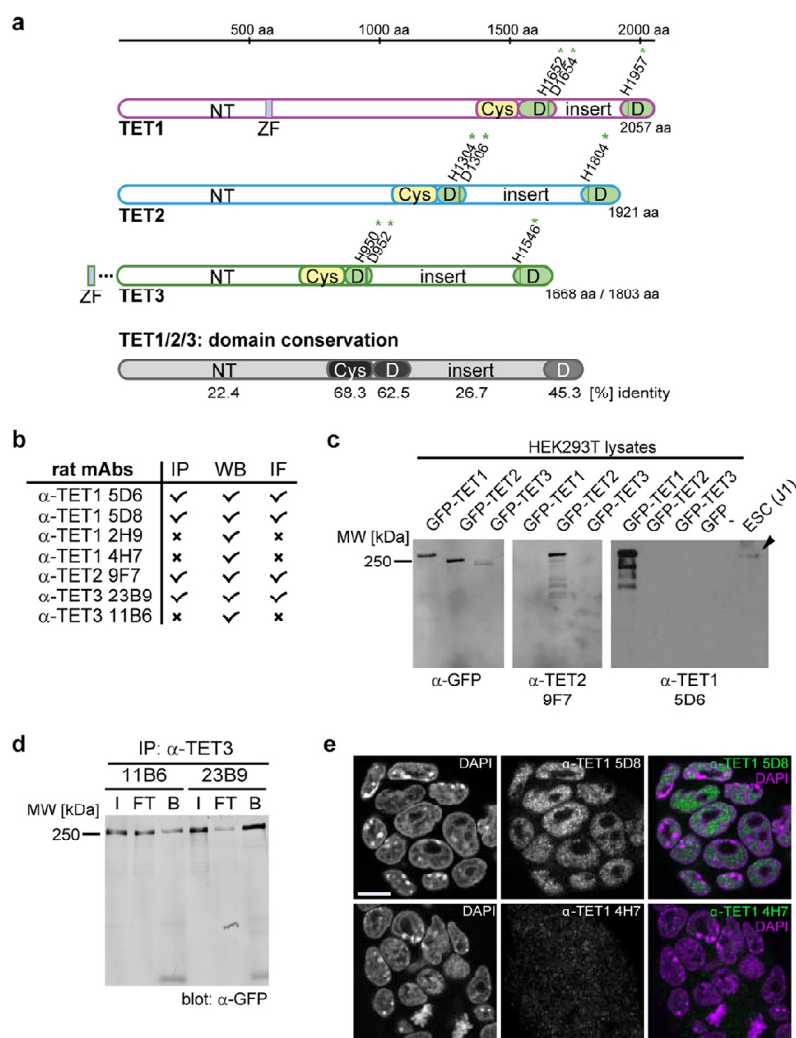
We generated antibodies against murine TET1, TET2, and TET3 using protein fragments derived from the insert region of the catalytic domains as antigens. The rat monoclonal antibodies were tested for their applicability in Western blotting, immunoprecipitation, and immunofluorescence (Fig. 1*b*). Fig. 1 (*c–e*) shows exemplary data from the antibody characterization process of selected clones. For antibody testing, GFP-tagged TET proteins were expressed in HEK293T cells and detected in the cell lysate by Western blotting using anti-TET antibodies and an anti-GFP antibody as a positive control (Fig. 1*c* and data not shown). For immunoprecipitation, antibodies were coupled to protein G beads, incubated with the cell lysates, and analyzed by anti-GFP Western blotting for efficient pull-down of the respective TET protein (Fig. 1*d* and data not shown). mESCs were used to test the suitability of the obtained antibodies for immunofluorescence. Antibodies preselected for specificity in Western blot analyses that showed a clear nuclear staining were judged as applicable in immunofluorescence (Fig. 1*e* and data not shown).

**TET Proteins Interact with and Are O-GlcNAcylated by OGT**—As a first step toward understanding the regulation of TET proteins, we screened for interaction partners in mESCs. Because TET1 and TET2 are constitutively expressed in mESCs, clones 5D6 (anti-TET1), 5D8 (anti-TET1), and 9F7 (anti-TET2) were used to pull down endogenous TET1 and TET2. Subsequent LC-MS/MS analysis revealed that both TET1 and TET2 interacted with the glycosyltransferase OGT. In accordance with this result, co-immunoprecipitation analysis of GFP-tagged TET1, TET2, and TET3 expressed in HEK293T cells shows high enrichment of OGT in the pull-down (Fig. 2*a*).

Having observed the interaction between TET proteins and OGT, we examined whether TET proteins are modified by OGT and screened for O-GlcNAcylation, the modification that is transferred to the OH group of serine or threonine residues of target proteins by OGT (38, 49). To this end, we specifically enriched GFP-tagged TET proteins coexpressed with either OGT or its catalytically inactive point mutant OGT<sup>mut</sup> with GFP-Trap and probed the subsequent Western blot with an anti-GlcNAc antibody. All three TET proteins were found to

## 2. Results

### Phosphorylation and O-GlcNAcylation of TET Proteins



**FIGURE 1. Generation of anti-TET monoclonal antibodies.** *a*, schematic representation of the domain architecture of the three murine TET proteins. The catalytic dioxygenase domain (D) is split in two parts, separated by a presumably unstructured low-complexity insert (8), and is N-terminally preceded by a cysteine-rich region (Cys). The Fe(II)-binding residues are marked with green asterisks. The N terminus (NT) of TET1 contains a CXXC-type zinc finger (ZF). TET3 exists in two isoforms, one with a zinc finger and one without (41). The mean percent identity of the single domains of TET1, TET2, and TET3 is represented by different shades of gray and was calculated with Clustal 2.1 (59). *aa*, amino acids. *b*, overview of the generated anti-TET monoclonal antibodies (*mAbs*) and their possible applications. *IP*, immunoprecipitation; *WB*, Western blotting; *IF*, immunofluorescence; *x*, antibody not suited for the indicated application. *c*, example of Western blot analysis of two anti-TET antibodies with an anti-GFP antibody as a positive control. The antibodies detected only their target protein, but not the other two TET proteins. The WT protein from mESC whole cell lysates was also detected specifically (black arrowhead). *d*, example of an immunoprecipitation experiment with the indicated anti-TET3 antibodies. Clone 23B9 efficiently precipitated TET3 compared with clone 11B6. Western blot analysis was performed with an anti-GFP antibody. *I*, input; *FT*, flow-through; *B*, bound. *e*, immunofluorescence staining of mESCs with anti-TET1 antibodies (clones 5D8 and 4H7) and DAPI as a DNA counterstain. Whereas clone 5D8 showed a clear nuclear pattern, clone 4H7 displayed only a weak and diffuse signal. Confocal imaging was performed with a Leica TCS SP5 confocal laser scanning microscope with a ×63/1.4 numerical aperture Plan-Apochromat oil immersion objective. Scale bar = 5 μm.

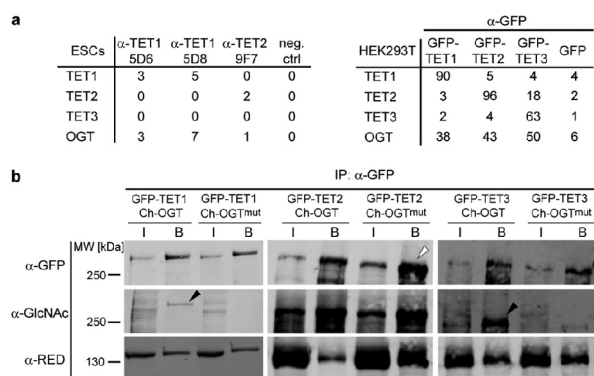
be increasingly O-GlcNAcylated upon the coexpression of catalytically active OGT (Fig. 2*b*).

**O-GlcNAcylation Reduces Phosphorylation of TET Proteins—** To identify OGT-dependent O-GlcNAcylation sites on TET proteins, we performed mass spectrometric analysis of semipurified proteins. We therefore expressed GFP-tagged TET1, TET2, and TET3 in HEK293T cells with either OGT or OGT<sup>mut</sup> or without interactor. After pulldown with GFP-Trap and stringent washing steps, the samples were analyzed by LC-MS/MS. An overall

sequence coverage of ~50% was achieved for TET1, ~60% for TET2, and ~65% for TET3 (supplemental Data S1 and Table S4). For data analysis, only sites were considered that were detected in at least two of three biological replicates. Fig. 3*a* shows an exemplary MS/MS spectrum of an O-GlcNAcylated TET1 peptide. Without coexpression of interactor, only a few residues on TET proteins were found to be O-GlcNAcylated at low site occupancy. Coexpression of OGT led to a strong increase in both the number of O-GlcNAcylation sites and site



## Phosphorylation and O-GlcNAcylation of TET Proteins



**FIGURE 2. TET proteins interact with OGT and are O-GlcNAcylated.** *a*, number of unique peptides detected in immunoprecipitation experiments followed by LC-MS/MS. *Left*, immunoprecipitation of endogenous TET1 or TET2 with the indicated antibodies. Protein G beads without antibody were used as a negative control (*neg. ctrl*). *Right*, immunoprecipitation of GFP-tagged TET1, TET2, or TET3 expressed in HEK293T cells. Pull-down of GFP served as a negative control. *b*, Western blot analysis of TET1, TET2, and TET3 specifically enriched with GFP-Trap. Upon coexpression of active OGT, the O-GlcNAcylation signal increased for TET1 and TET3 (*black arrowheads*) compared with coexpression of catalytically inactive OGT<sup>mut</sup>. For TET2, protein levels in the OGT<sup>mut</sup> samples were higher (*white arrowhead*), whereas the O-GlcNAc signal remained constant, suggesting a higher proportion of O-GlcNAcylated TET2 in the OGT sample. Interaction between TET proteins and OGT was independent of OGT activity. Anti-RED antibody (60) detected the coexpressed Cherry (*Ch*)-tagged OGT. *IP*, immunoprecipitation; *I*, input; *B*, bound.

occupancy for TET2 and TET3. The difference in the number of O-GlcNAc sites was either due to *de novo* modification by OGT or because the site occupancy without OGT coexpression was below the detection limit. For TET1, however, the O-GlcNAc pattern was relatively heterogeneous, and only a few O-GlcNAc sites could be detected. This heterogeneity is also illustrated by the fact that residues 1327 and 327, which were O-GlcNAcylated in the TET1 samples, were modified only in one of three replicates in the TET1/OGT samples. Although Cherry-OGT<sup>mut</sup> is supposed to be catalytically inactive, coexpression led to a small increase in O-GlcNAcylation and represented a distinct state from basal levels (Fig. 4 and Tables 1–3).

Because O-GlcNAcylation occurs at serine or threonine residues of the target protein, we also screened for another PTM that can occur at these amino acids, namely phosphorylation. Interestingly, high phosphorylation of TET1, TET2, and, to a lesser extent, TET3 was observed. Phosphorylation of all TET proteins decreased significantly upon coexpression of OGT regarding both site occupancy and the number of detected phosphorylation sites (Fig. 4 and Tables 1–3). An example of a MS/MS spectrum of a phosphorylated TET1 peptide is shown in Fig. 3*b*. The MS/MS spectra of all modified TET peptides are provided in [supplemental Data S2 and S3](#).

**PTMs Occur Mostly at the N terminus and in the Low-complexity Insert of TET Proteins**—To date, the domains of TET proteins are largely uncharacterized, except for the conserved catalytic dioxygenase domain and the CXXC-type zinc finger at the N terminus of TET1 (8, 40, 48). Mapping the detected O-GlcNAc and phosphorylation sites to the TET protein sequence revealed that mostly the N terminus and low-complexity insert, which separates the two parts of the dioxy-

genase domain, were subjected to PTMs (Fig. 5). Remarkably, O-GlcNAcylation and phosphorylation rarely occurred at the exact same residue, although O-GlcNAcylation suppressed phosphorylation. Furthermore, the three TET proteins had different modification patterns: whereas TET1 was modified mostly at the N terminus and the very C-terminal part and was hardly glycosylated, TET2 and TET3 showed strong O-GlcNAcylation in the low-complexity insert region. The first 350 amino acids of TET3 remained free of PTMs. The observed pattern is not due to differences in sequence coverage, as the detected peptides are homogeneously distributed over the whole protein sequence ([supplemental Data S1](#)).

Interestingly, some of the modifications were detected on the same peptides, indicating that they occurred together at the same molecule. For example, TET2 Ser-23 phosphorylation could be found with Ser-15 phosphorylation, and Ser-376 phosphorylation occurred only when Ser-374 was O-GlcNAcylated, but not when it was phosphorylated (Table 2). For TET3, a variety of PTM combinations could be observed for residues 360–368 and 1071–1077. Phosphorylation at Ser-362, for example, existed either alone or in combination with Ser-360 O-GlcNAcylation and Ser-368 phosphorylation. Phosphorylation of Ser-362 also co-occurred with O-GlcNAcylation of Ser-361. If Ser-362 was O-GlcNAcylated, however, no further modifications on this peptide were observed (Fig. 4 and Table 3). Apparently, some residues such as TET3 Ser-362 serve as O-GlcNAcylation/phosphorylation switches that can either promote or suppress neighboring PTMs. These data indicate a strong cross-talk between O-GlcNAcylation and phosphorylation at different residues. On the other hand, modifications on TET1 appeared more isolated, and no peptide bearing more than one modification was detected (Table 1). In summary, we detected many interdependent modification sites on TET proteins, suggesting that TET1, TET2, and TET3 are dynamically regulated by PTMs.

## DISCUSSION

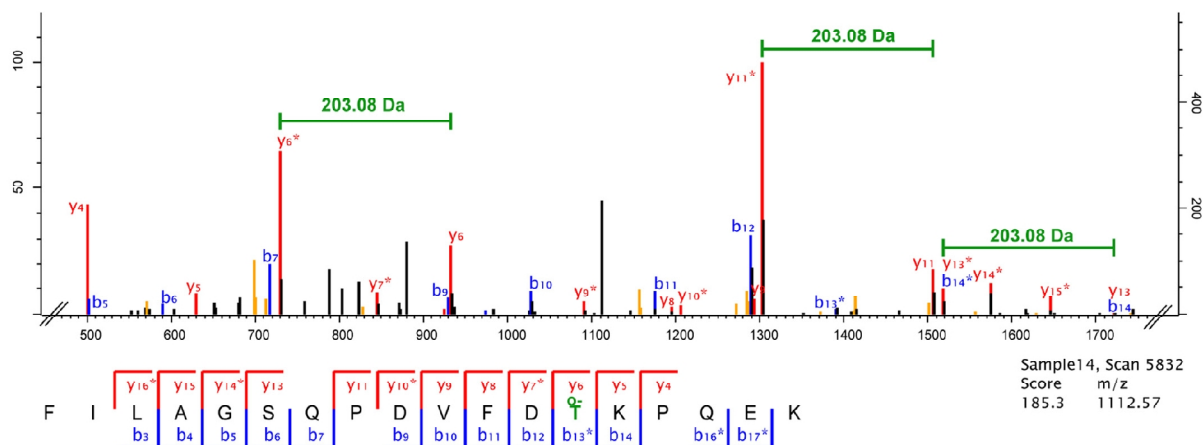
Because oxidation of 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine by TET proteins represents a potential mechanism for active DNA demethylation in higher vertebrates (3–5), these proteins are intensively investigated. Here, we provide evidence that all three TET proteins are subject to O-GlcNAcylation through OGT. This finding is in accordance with previous studies showing that TET1 and TET2 interact with OGT in embryonic stem cells and are O-GlcNAcylated (33, 34). TET3 has also been described to associate with OGT (32, 35) and to alter its subcellular localization dependent on glucose metabolism and O-GlcNAcylation (35). Not only does OGT directly modify TET proteins, but the interaction also promotes histone modifications such as H3K4me3 and H2BS112GlcNAc (31, 36). TET1 has been shown to associate with the repressive SIN3A complex (50), and TET2 and TET3 have been shown to associate with the SET1/COMPASS complex (31).

We have shown that, by default, TET proteins are phosphorylated. Basal O-GlcNAc levels are low but increase upon OGT expression. Simultaneously, the phosphorylation levels decrease. This finding identifies regulation of the phosphorylation signal as a

## 2. Results

### Phosphorylation and O-GlcNAcylation of TET Proteins

#### a TET1 T327-GlcNAc



#### b TET1 S320-ph

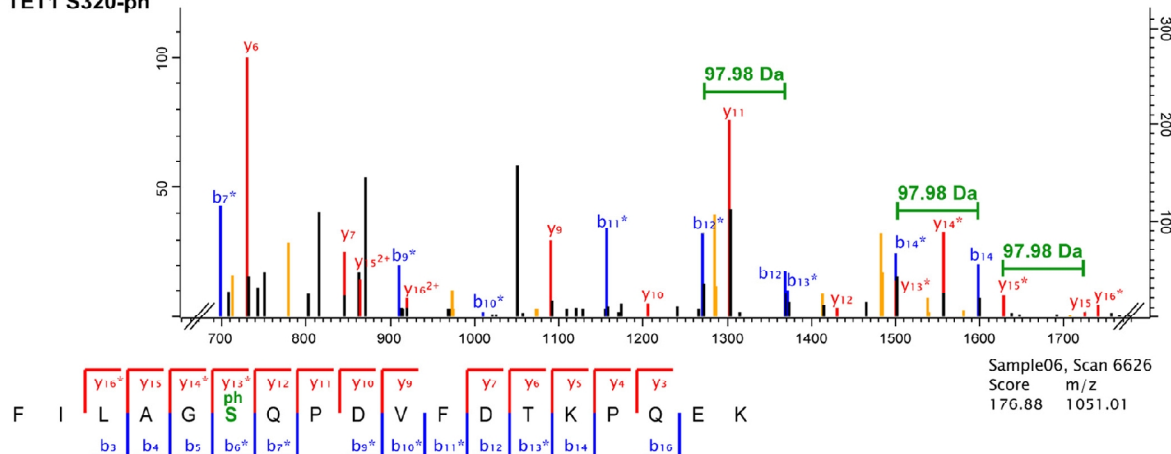


FIGURE 3. Exemplary MS/MS spectra of modified TET1 peptides. *a*, MS/MS spectrum of a TET1 peptide modified with O-GlcNAc (*o*-) at the threonine residue. O-GlcNAcylation is characterized by a neutral loss of 203.8 Da as indicated. *y* ions are depicted in red, and *b* ions are depicted in blue. Labeling of neutral losses of H<sub>2</sub>O or NH<sub>3</sub> (orange peaks) has been removed for clarity. Fully annotated spectra are provided in supplemental Data S2 and S3. *b*, MS/MS spectrum of the same TET1 peptide phosphorylated (*ph*) at the serine residue. Phosphorylated ions show a neutral loss of 97.98 Da as indicated. *y* ions are depicted in red, *b* ions are depicted in blue. Labeling of neutral losses of H<sub>2</sub>O or NH<sub>3</sub> (orange peaks) has been removed for clarity. Fully annotated spectra are provided in supplemental Data S2 and S3.

novel function for TET O-GlcNAcylation. Interestingly, the underlying mechanism of this observation seems not to be direct competition for the serine or threonine residue that is to be modified, but rather proximal site competition as neighboring residues are interdependent (51). O-GlcNAcylation and phosphorylation of TET proteins occur at distinct amino acids, and several modifications of the same type often appear in close proximity in “modification islands,” *e.g.* O-GlcNAcylation at Ser-1252/Ser-1256/Ser-1263 of TET3 or phosphorylation at Ser-15/Ser-23/Ser-39 of TET2. It is important to note that only a few and more isolated O-GlcNAcylation sites are detected on TET1 compared with TET2 and TET3 and that glycosylation of TET1 is less conserved within biological replicates. We also did not observe O-GlcNAcylation of TET1 at Thr-535, which has been described previously as a major TET1 glycosylation site (33, 52). O-GlcNAcylation of TET1 seems to be very dynamic. This

hypothesis is also supported by the fact that Myers *et al.* (52) detected TET1 Thr-535 O-GlcNAcylation in only one of three replicates, similar to our observation of heterogeneous TET1 glycosylation patterns.

To distinguish between mere interaction of TET proteins with OGT and catalytic activity of OGT on TET proteins, we used a catalytically inactive point mutant of OGT as a control. Interestingly, O-GlcNAcylation of TET proteins was slightly increased by OGT<sup>mut</sup>. This might be due to residual activity of the mutant (53) or, more likely, to recruitment of endogenous active OGT via trimerization of the tetratricopeptide repeat domain (54). Nevertheless, this supposed heterotrimer seems to target the same residues, as 91% of all detected O-GlcNAc sites in the OGT<sup>mut</sup> samples were also modified in the OGT samples. Regarding phosphorylation, coexpression of OGT<sup>mut</sup> also represents an intermediate state between the basal state,

## Phosphorylation and O-GlcNAcylation of TET Proteins

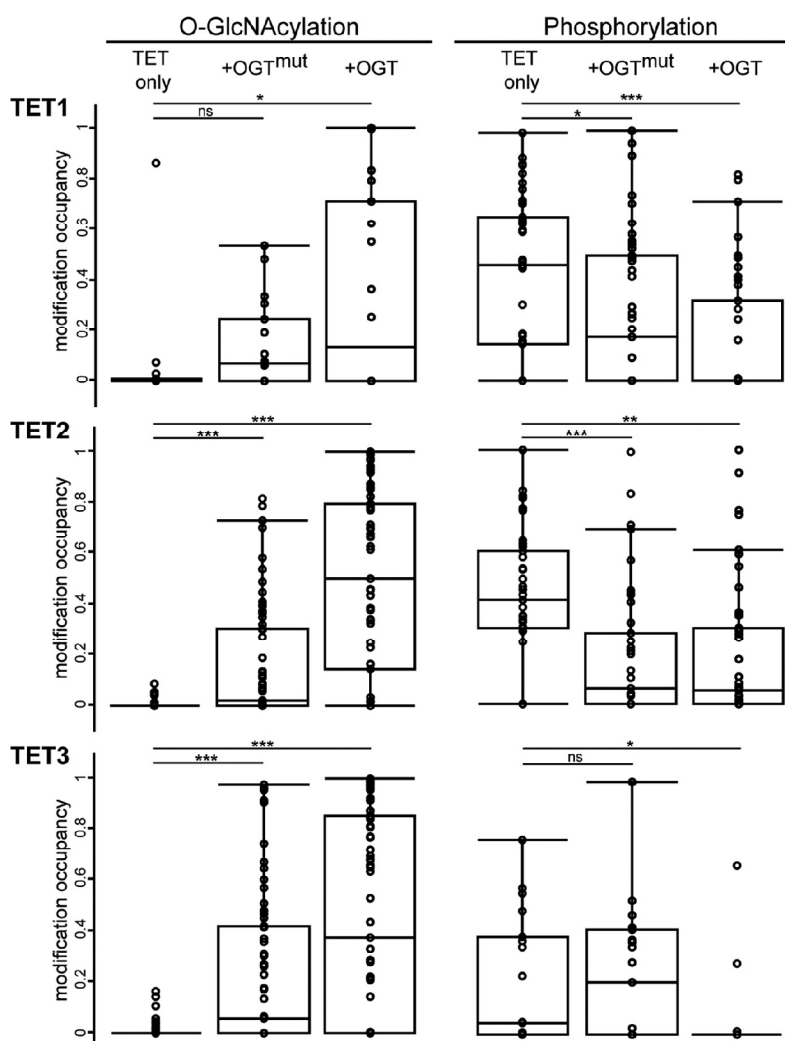


FIGURE 4. **TET phosphorylation is reduced upon O-GlcNAcylation.** Box plots depict the distribution of O-GlcNAc and phosphorylation occupancy in the three conditions: expression of TET protein only, coexpression of OGT<sup>mut</sup>, and coexpression of OGT. Missing values have been substituted with an occupancy of 0.005, with 0.001 being the lowest measured occupancy. Mean occupancies of single sites are provided in Tables 1–3. \*,  $p < 0.05$  (Student's *t* test); \*\*,  $p < 0.01$ ; \*\*\*,  $p < 0.001$ ; ns, not significant.

*i.e.* only TET expression, and coexpression of active OGT. 74% of all detected phosphorylation sites in the basal state also appeared during coexpression of OGT<sup>mut</sup>, arguing against steric hindrance of the kinase by the inactive enzyme as a mechanism for reduced phosphorylation.

In this study, we investigated TET protein PTMs dependent on OGT levels. The observed effect that O-GlcNAcylation of TET proteins reduces phosphorylation is of particular interest because protein O-GlcNAc levels are influenced by a variety of factors, such as different subcellular localization of OGT and nutrient availability, and seem to be tightly regulated (38). For example, O-GlcNAcylation of TET3 can be enhanced when cells are cultured in high-glucose medium, leading to nuclear export of TET3 (35). Furthermore, not only OGT activity but also OGT expression levels are tightly controlled in living cells. During chondrocyte differentiation, for example, OGT is up-

regulated upon insulin stimulation (55). Moreover, the *Ogt* gene is located on the X chromosome and is subjected to dosage compensation through X chromosome inactivation (56). OGT-dependent dephosphorylation represents a novel mechanism by which TET proteins could be regulated in response to changing environmental conditions or during differentiation.

Interestingly, some residues remain stably phosphorylated even at high OGT levels. For TET2, they appear in close proximity to each other and just N-terminal of the cysteine-rich region. This persistence of phosphorylation suggests an important OGT-independent regulatory role of these residues that is of interest for future studies. Nevertheless, the majority of phosphorylation sites are reduced in occupancy upon O-GlcNAcylation. We thus observe two different types of phosphorylation: dependent on and independent of O-GlcNAcylation.

Phosphorylation and O-GlcNAcylation of TET Proteins

TABLE 1  
Detected modified peptides of TET1

(ph), phosphorylated; (o), O-GlcNAcylated; (ox), oxidized. Localization probability was calculated with MaxQuant software (44). Residue numbering refers to the murine protein sequences specified in supplemental Data S1. The arithmetic mean  $\pm$  S.D. of the occupancy is depicted for each data set. ND, not detected.

Modified amino acid	Localization probability	Modified sequence	Mean TET1	Mean TET1 + OGT <sup>mut</sup>	Mean TET1 + OGT
160	1.00	H...ATVS(ph)PGTENGEQNR	0.15 $\pm$ ND	0.98 $\pm$ ND	0.34 $\pm$ 0.08
177	1.00	CLVEGES(ph)QEITQSCPVEER	0.63 $\pm$ 0.01	0.32 $\pm$ 0.15	0.48 $\pm$ ND
253	0.85	NT(o)SNQLADLSSQVESIK	ND	0.07 $\pm$ 0.01	0.37 $\pm$ ND
270	0.84	LS(o)DPSPNPTGSDHNGFPDSSFR	ND	0.08 $\pm$ 0.02	0.67 $\pm$ 0.06
320	1.00	FILAGS(ph)QPDVFDTKPQEK	0.60 $\pm$ 0.16	0.44 $\pm$ 0.03	0.30 $\pm$ 0.20
327	1.00	FILAGSQPDVFDT(o)KPQEK <sup>a</sup>	0.32 $\pm$ 0.47	0.36 $\pm$ 0.15	0.55 $\pm$ ND
556	0.81	A...STSS(ph)PPCNSTPPMVER	0.23 $\pm$ 0.10	ND	ND
561	0.89	A...STSSPPCNS(ph)TPPM(ox)VER	0.20 $\pm$ 0.09	0.88 $\pm$ ND	ND
734	0.98	QQTNPS(ph)PTFAQTIR	0.44 $\pm$ ND	0.46 $\pm$ 0.33	0.32 $\pm$ ND
736	0.96	QQTNPSPT(ph)FAQTIR	0.67 $\pm$ 0.06	ND	ND
794	0.77	DAM(ox)SVTTS(o)GGECDHKLK	ND	0.48 $\pm$ ND	1.00 $\pm$ 0.00
854	1.00	DGS(ph)PVQPSLLSLMK	0.73 $\pm$ 0.13	0.54 $\pm$ 0.07	0.25 $\pm$ 0.34
892	0.70	L...SESSS(ph)PSKPEK	0.51 $\pm$ 0.48	0.27 $\pm$ 0.03	0.79 $\pm$ ND
950	1.00	S(ph)PDSFATNQALIK <sup>b</sup>	0.68 $\pm$ 0.26	0.72 $\pm$ 0.20	0.49 $\pm$ 0.11
969	0.74	SQGYPS(ph)PT...	0.61 $\pm$ 0.03	ND	ND
1327	0.66	REAQT(o)SSN...K <sup>a</sup>	0.01 $\pm$ 0.00	ND	0.79 $\pm$ ND
1964	0.89	ELHATTSLSRS(ph)PK	0.33 $\pm$ 0.21	0.17 $\pm$ ND	0.47 $\pm$ 0.33
2016	1.00	PADRECPDVS(ph)PEANLSHQIPSR	0.68 $\pm$ 0.21	0.37 $\pm$ 0.18	0.81 $\pm$ ND
2016	0.56	PADRECPDVS(o)PEANLSHQIPSR	ND	0.26 $\pm$ 0.10	0.55 $\pm$ 0.41
2042	0.99	DNVVTVS(ph)PYSLTHVAGPYNR	0.73 $\pm$ 0.12	ND	0.38 $\pm$ ND

<sup>a</sup> Basal O-GlcNAc sites.  
<sup>b</sup> Persistent phosphorylation sites.

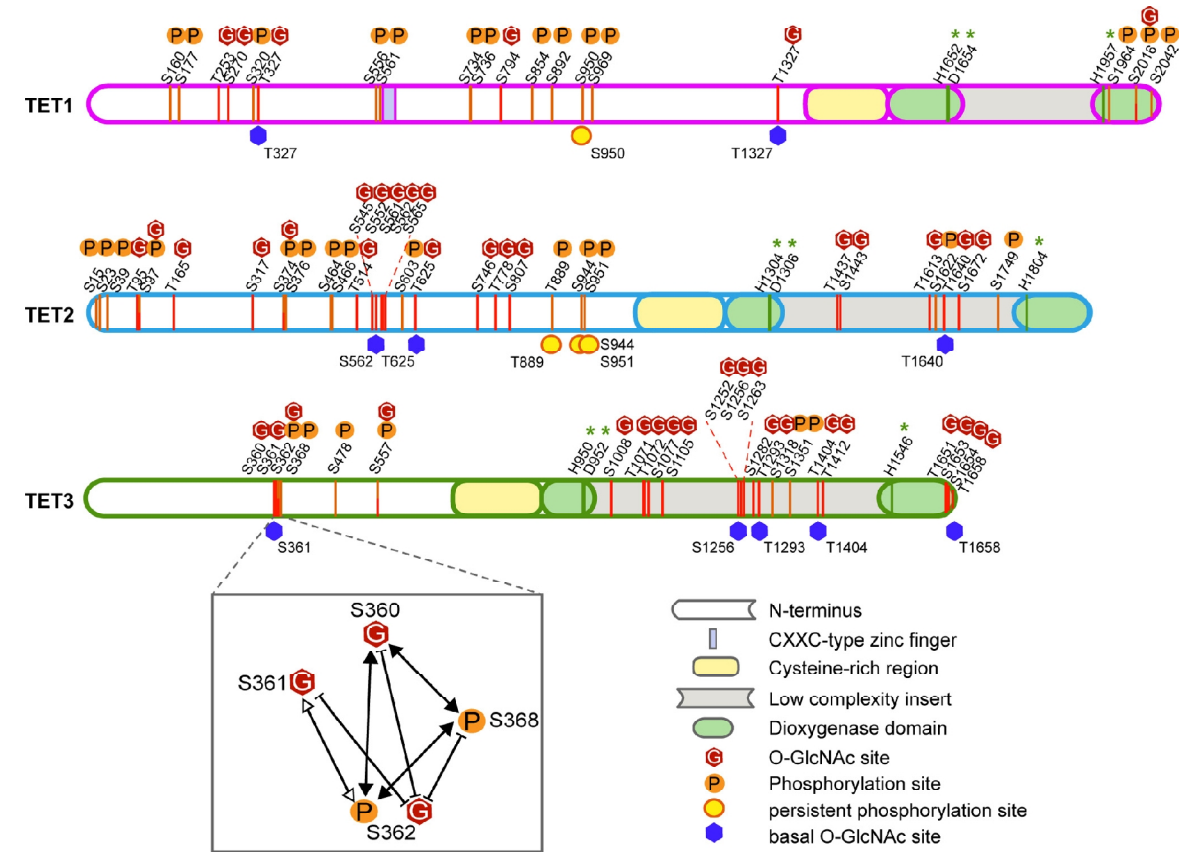


FIGURE 5. N termini and insert regions of TET proteins are densely modified. Shown are a schematic and scaled mapping of all TET phosphorylation and O-GlcNAcylation sites in the protein sequence. Modifications are found mostly in the N terminus and insert region and rarely occur at the same residue. Residue numbering refers to the murine protein sequences specified in supplemental Data S1. Green asterisks indicate catalytic Fe(II)-binding residues. Basal O-GlcNAc sites occur without any coexpression of OGT or OGT<sup>mut</sup>; persistent phosphorylation sites show high occupancy despite an increase in O-GlcNAcylation. An example of the PTM cross-talk on TET proteins is shown for TET3 Ser-360/Ser-361/Ser-362/Ser-368. White arrowheads, two co-occurring modifications; black arrowheads, three co-occurring modifications; blunt arrows, mutual exclusivity.

## Phosphorylation and O-GlcNAcylation of TET Proteins

**TABLE 2****Detected modified peptides of TET2**

(ph), phosphorylated; (o), O-GlcNAcylated; (ox), oxidized. Multiple modifications occurring on one peptide are shown in boldface. Localization probability was calculated with the MaxQuant software (44). Residue numbering refers to the murine protein sequences specified in [supplemental Data S1](#). The arithmetic mean  $\pm$  S.D. of the occupancy is depicted for each data set. ND, not detected.

Modified amino acid	Localization probability	Modified sequence	Mean TET2	Mean TET2 + OGT <sup>mut</sup>	Mean TET2 + OGT
15	1.00	TTHAEGTRL(ph)PFLIAPPS...K	0.57 $\pm$ 0.06	0.20 $\pm$ 0.01	0.01 $\pm$ 0.00
23	1.00	T...LS(ph)PFLIAPPS(ph)PIS...K	0.66 $\pm$ 0.09	0.32 $\pm$ 0.17	0.32 $\pm$ 0.04
39	0.98	LQNGS(ph)PLAERPHPEVNGDTK	0.45 $\pm$ 0.10	ND	ND
95	0.98	RT(o)VS(o)EPSLSGLHPNK	ND	0.06 $\pm$ 0.00	0.26 $\pm$ 0.17
97	1.00	TVS(ph)EPSLSGLHPNK	0.53 $\pm$ 0.26	0.49 $\pm$ 0.30	0.29 $\pm$ 0.02
97	0.97	RT(o)VS(o)EPSLSGLHPNK	ND	0.01 $\pm$ ND	0.29 $\pm$ 0.06
165	1.00	S...TSTTQESSGADAFPT(o)R	ND	0.74 $\pm$ 0.06	0.98 $\pm$ 0.02
317	0.98	SALDIGPS(o)RAENK	ND	ND	0.48 $\pm$ 0.03
374	0.82	DS(ph)ISPTTVTPPSQSLAPR	ND	ND	0.34 $\pm$ 0.49
374	0.99	DS(o)IS(ph)PTTVTPPSQSLAPR	ND	0.19 $\pm$ 0.22	0.49 $\pm$ 0.37
376	0.99	DS(o)IS(ph)PTTVTPPSQSLAPR	1.00 $\pm$ ND	0.44 $\pm$ 0.38	0.30 $\pm$ 0.20
464	1.00	T...LPEQHQNDCGS(ph)PS(ph)PEK	0.79 $\pm$ 0.03	ND	ND
466	1.00	T...LPEQHQNDCGS(ph)PS(ph)PEK	0.79 $\pm$ 0.03	ND	ND
514	0.89	QT(o)QGSVQAAPGWIELK	ND	0.09 $\pm$ 0.03	0.59 $\pm$ 0.23
545	0.94	DIS(o)LHSLVHSQT...M(ox)SSK	ND	0.46 $\pm$ 0.10	0.78 $\pm$ 0.02
552	0.87	DIS(o)LHSLVHS(o)QT...MSSK	ND	0.13 $\pm$ ND	0.76 $\pm$ 0.20
561	0.95	DIS...VNQMS(o)S(o)K	ND	0.07 $\pm$ 0.00	0.79 $\pm$ 0.17
562	0.97	DIS...VNQMS(o)S(o)K <sup>a</sup>	0.01 $\pm$ 0.01	0.42 $\pm$ 0.06	0.80 $\pm$ 0.17
565	0.98	QS(o)TGNVNM(ox)PGGFQR	ND	ND	0.41 $\pm$ 0.04
603	1.00	AQMYQVQVNVQGPS(ph)PG...K	0.41 $\pm$ 0.17	0.06 $\pm$ ND	0.14 $\pm$ 0.05
625	0.96	ALYQECIPRT(o)DPSS...R <sup>a</sup>	0.05 $\pm$ 0.01	0.73 $\pm$ 0.13	0.98 $\pm$ 0.02
746	0.98	VEESFCVGNQYS(o)K	ND	0.23 $\pm$ 0.16	0.83 $\pm$ 0.06
778	0.92	ILT(o)PNSSNLQILPSNDTHPACER	0.09 $\pm$ ND	0.31 $\pm$ 0.01	0.64 $\pm$ 0.05
807	1.00	EQALHPVGS(o)K	ND	0.01 $\pm$ ND	0.58 $\pm$ 0.12
889	1.00	ALPVPEQGGSTQTQT(ph)PPQK <sup>b</sup>	0.57 $\pm$ 0.23	0.78 $\pm$ 0.30	0.57 $\pm$ 0.05
944	1.00	YPLS(ph)PPQENMSSR <sup>b</sup>	0.43 $\pm$ 0.14	0.46 $\pm$ 0.22	0.67 $\pm$ 0.11
951	0.97	PSSYRYPLSPQENMS(ph)SR <sup>b</sup>	0.24 $\pm$ ND	0.32 $\pm$ ND	0.51 $\pm$ 0.68
1437	0.63	QM(ox)T(o)AQPQLS...R	ND	ND	0.67 $\pm$ 0.06
1443	0.98	QMTAQPQLS(o)GPVIR	ND	0.05 $\pm$ 0.05	0.52 $\pm$ 0.42
1613	0.87	D...PPIHT(o)LHQQTFGDSPSK	ND	ND	0.09 $\pm$ 0.10
1622	0.74	Y...TLHQQTFGDS(ph)PSK	0.45 $\pm$ 0.15	0.07 $\pm$ 0.04	0.76 $\pm$ ND
1640	0.76	DAFT(o)TNSTLKP...K <sup>a</sup>	0.05 $\pm$ 0.01	0.50 $\pm$ 0.20	0.84 $\pm$ 0.03
1672	1.00	M(ox)DSHEM(ox)GAAS(o)R	ND	ND	0.93 $\pm$ 0.01
1749	1.00	TASAQELLYSLTGSS(ph)QEK	0.31 $\pm$ 0.02	0.07 $\pm$ 0.05	0.27 $\pm$ 0.02

<sup>a</sup> Basal O-GlcNAc sites.

<sup>b</sup> Persistent phosphorylation sites.

**TABLE 3****Detected modified peptides of TET3**

(ph), phosphorylated; (o), O-GlcNAcylated; (ox), oxidized. Multiple modifications occurring on one peptide are shown in boldface. Localization probability was calculated with MaxQuant software (44). Residue numbering refers to the murine protein sequences specified in [supplemental Data S1](#). The arithmetic mean  $\pm$  S.D. of the occupancy is depicted for each data set. ND, not detected.

Modified amino acid	Localization probability	Modified sequence	Mean TET3	Mean TET3 + OGT <sup>mut</sup>	Mean TET3 + OGT
360	0.93	VEAPS(o)SS(ph)PAPVPS(ph)PISQR	ND	0.10 $\pm$ 0.07	0.91 $\pm$ 0.09
361	0.79	VEAPSS(o)S(ph)PAPVPSISQR <sup>a</sup>	0.02 $\pm$ 0.01	0.52 $\pm$ 0.13	ND
362	1.00	VEAPSSS(ph)PAPVPSISQR	0.03 $\pm$ 0.03	0.38 $\pm$ 0.03	0.01 $\pm$ ND
362	0.67	VEAPSSS(o)PAPVPSISQR	ND	0.96 $\pm$ 0.02	ND
368	1.00	VEAPS(o)SS(ph)PAPVPS(ph)PISQR	0.34 $\pm$ ND	0.55 $\pm$ 0.40	0.28 $\pm$ ND
478	1.00	S(ph)RDM(ox)QPLFLPVR	0.46 $\pm$ 0.13	0.38 $\pm$ 0.13	0.66 $\pm$ ND
557	0.83	S(ph)PSPM(ox)VALQSGST...R	0.23 $\pm$ ND	0.22 $\pm$ 0.27	ND
557	0.76	S(o)PSPM(ox)VALQSGST...R	ND	ND	0.44 $\pm$ 0.30
1008	0.83	VS(o)SGAIQVLTAFPR	ND	0.91 $\pm$ 0.01	0.36 $\pm$ 0.51
1071	0.97	QEALAGVTT(o)T(o)DPGLSLK	ND	ND	0.96 $\pm$ 0.01
1072	0.89	QEALAGVTT(o)T(o)DPGLSLK	ND	ND	0.96 $\pm$ 0.01
1077	0.99	QEALAGVTT(o)DPGLS(o)LK	ND	0.50 $\pm$ 0.40	0.53 $\pm$ 0.31
1105	0.89	YS(o)GNVSVESYVLGS...R	ND	0.40 $\pm$ 0.07	0.73 $\pm$ 0.11
1252	0.94	VPQLHPAS(o)RDPSPFAQSSSCYNR	ND	0.42 $\pm$ ND	0.95 $\pm$ 0.03
1256	0.62	VPQLHPASRDPS(o)PFAQSSSCYNR <sup>a</sup>	0.04 $\pm$ 0.01	ND	0.98 $\pm$ 0.01
1263	0.84	VPQLHPASRDPSFAQSSS(o)CYNR	ND	0.48 $\pm$ ND	0.98 $\pm$ 0.03
1282	0.88	QEPIDPLTQAES(o)IPR	ND	0.30 $\pm$ 0.06	0.91 $\pm$ 0.09
1293	1.00	T(o)PLPEAS...SGGPSMSPK <sup>a</sup>	0.01 $\pm$ 0.00	0.53 $\pm$ 0.06	0.99 $\pm$ 0.01
1318	1.00	TPLPEAS...SGGPSM(ox)S(ph)PK	0.43 $\pm$ 0.07	0.00 $\pm$ ND	ND
1351	0.61	LNSFGAS(ph)CLTPSHFPES...R	0.45 $\pm$ 0.39	ND	ND
1404	0.76	FGNGTSALTGPSLT(o)EK <sup>a</sup>	0.02 $\pm$ 0.02	ND	0.74 $\pm$ 0.15
1412	1.00	PWGM(ox)GT(o)GDFNPALK	ND	0.06 $\pm$ 0.00	0.65 $\pm$ 0.14
1651	0.72	Q...SAVT(o)VSSYAYTK	ND	ND	0.18 $\pm$ 0.05
1653	0.71	Q...SAVTVS(o)SYAYTK	ND	0.43 $\pm$ 0.19	0.73 $\pm$ 0.05
1654	0.77	Q...SAVTVS(o)S(o)YAYTK	ND	ND	0.24 $\pm$ 0.05
1658	0.99	G...TDSAVTVSSYAYT(o)K <sup>a</sup>	0.14 $\pm$ 0.03	0.12 $\pm$ 0.05	0.75 $\pm$ 0.14

<sup>a</sup> Basal O-GlcNAc sites.



## Phosphorylation and O-GlcNAcylation of TET Proteins

The hypothesis of interdependence of PTMs on TET proteins is further strengthened by the fact that some modifications are detected on the same peptides in stable combinations, whereas others occur as stand-alone modifications. Certain residues appear to be O-GlcNAcylation/phosphorylation switches that influence the PTM pattern on the neighboring amino acids. The observed cross-talk of modifications enables a variety of potential regulatory mechanisms that could fine-tune TET activity dependent on different environmental conditions such as nutrient availability.

To date, the domain architecture and three-dimensional structure of TET proteins are only poorly understood. The catalytic domain is highly conserved and homologous to other types of Fe(II)- and 2-oxoglutarate-dependent dioxygenases that act on nucleic acids (8, 57). Recently, the crystal structure of the catalytic region of TET2 has provided insights into the reaction mechanism (48). However, the large N terminus and low-complexity insert, which is characteristic for TET proteins, remain poorly understood in terms of both structure and function. So far, no homologous domains have been described, except for the CXXC-type zinc finger at the N terminus, and the insert region is predicted to be largely unstructured (8). In this study, we have shown that these two regions are subject to many dynamic PTMs. For TET1 and TET3, a few modification sites are also found at the very C terminus of the proteins, but the N terminus and insert region are the major targets of O-GlcNAcylation and phosphorylation. In general, the lower the conservation of one region, the more modification sites are detected. The selective modification of these regions might contribute to the regulation of TET protein activity, stability, or targeting. TET1, TET2, and TET3 have been described to co-localize with OGT at transcription start sites and influence gene expression (31, 34). The different modifications described in this study might alter binding of TET interaction partners and thus provide a possible explanation for the observed dual role in transcription activation and repression (58).

In summary, we have provided the first systematic mapping of O-GlcNAcylation and phosphorylation sites on TET proteins at amino acid resolution. The distribution of these PTMs and the described cross-talk provide new perspectives on the regulatory role of the so far poorly characterized non-catalytic domains: the N terminus and low-complexity insert region. The observed O-GlcNAcylation and phosphorylation are linked to metabolic conditions and thus provide a possible mechanism of TET protein regulation in response to external stimuli.

## REFERENCES

- Goll, M. G., and Bestor, T. H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514
- Suzuki, M. M., and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9**, 465–476
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303
- He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., and Xu, G. L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307
- Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C. A., Hagemeyer, C., and Carell, T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012
- Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S. K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotlarova, O., Schuermann, D., Michalakakis, S., Kosmatchev, O., Schiesser, S., Steigenberger, B., Raddaoui, N., Kashiwazaki, G., Müller, U., Spruijt, C. G., Vermeulen, M., Leonhardt, H., Schär, P., Müller, M., and Carell, T. (2014) Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581
- Iyer, L. M., Tahiliani, M., Rao, A., and Aravind, L. (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**, 1698–1710
- Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F., and Leonhardt, H. (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181
- Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., Shi, L., He, X., Jin, S. G., Iqbal, K., Shi, Y. G., Deng, Z., Szabó, P. E., Pfeifer, G. P., Li, J., and Xu, G. L. (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610
- Kim, M., Park, Y. K., Kang, T. W., Lee, S. H., Rhee, Y. H., Park, J. L., Kim, H. J., Lee, D., Lee, D., Kim, S. Y., and Kim, Y. S. (2014) Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. *Hum. Mol. Genet.* **23**, 657–667
- Langlois, T., da Costa Reis Monte-Mor, B., Lenglet, G., Droin, N., Marty, C., Le Couédic, J. P., Almire, C., Auger, N., Mercher, T., Delhommeau, F., Christensen, J., Helin, K., Debili, N., Fuks, F., Bernard, O. A., Solary, E., Vainchenker, W., and Plo, I. (2014) TET2 deficiency inhibits mesoderm and hematopoietic differentiation in human embryonic stem cells. *Stem Cells* **32**, 2084–2097
- Rudenko, A., Dawlaty, M. M., Seo, J., Cheng, A. W., Meng, J., Le, T., Faull, K. F., Jaenisch, R., and Tsai, L. H. (2013) Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109–1122
- Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., Zepeda-Martinez, J. A., Pape, U. J., Jacobsen, S. E., Peters, B., and Rao, A. (2014) Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1361–1366
- Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* **2**, 241
- Hahn, M. A., Qiu, R., Wu, X., Li, A. X., Zhang, H., Wang, J., Jui, J., Jin, S. G., Jiang, Y., Pfeifer, G. P., and Lu, Q. (2013) Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis. *Cell Rep.* **3**, 291–300
- Kriaucionis, S., and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930
- Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J. G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., Diao, J., Wu, F., He, H. H., Cui, Q., Clark, E., Ma, C., Barbara, A., Veenstra, G. J., Xu, G., Kaiser, U. B., Liu, X. S., Sugrue, S. P., He, X., Min, J., Kato, Y., and Shi, Y. G. (2012) Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and neural development. *Cell* **151**, 1200–1213
- Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R., Huberman, K., Thomas, S., Dolgalev, I., Heguy, A., Paietta, E., Le Beau, M. M., Beran, M., Tallman, M. S., Ebert, B. L., Kantarjian, H. M., Stone, R. M., Gilliland, D. G., Crispino, J. D., and Levine, R. L. (2009) Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114**, 144–147
- Konstandin, N., Bultmann, S., Szwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P. M., Schneider, S., Spiekermann, K., Leonhardt, H., and Bohlander, S. K. (2011) Genomic 5-hy-

## Phosphorylation and O-GlcNAcylation of TET Proteins

- droxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia* **25**, 1649–1652
21. Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J. P., Robert, F., Alberdi, A., Lécluse, Y., Plo, I., Dreyfus, F. J., Marzac, C., Casadevall, N., Lacombe, C., Romana, S. P., Dessen, P., Soulier, J., Viguié, F., Fontenay, M., Vainchenker, W., and Bernard, O. A. (2009) Mutation in TET2 in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301
  22. Ko, M., Huang, Y., Jankowska, A. M., Pape, U. J., Tahiliani, M., Bandukwala, H. S., An, J., Lamperti, E. D., Koh, K. P., Ganetzky, R., Liu, X. S., Aravind, L., Agarwal, S., Maciejewski, J. P., and Rao, A. (2010) Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843
  23. Dang, L., Jin, S., and Su, S. M. (2010) IDH mutations in glioma and acute myeloid leukemia. *Trends Mol. Med.* **16**, 387–397
  24. Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., Fantin, V. R., Jang, H. G., Jin, S., Keenan, M. C., Marks, K. M., Prins, R. M., Ward, P. S., Yen, K. E., Liao, L. M., Rabinowitz, J. D., Cantley, L. C., Thompson, C. B., Vander Heiden, M. G., and Su, S. M. (2009) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744
  25. Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S. H., Ito, S., Yang, C., Wang, P., Xiao, M. T., Liu, L. X., Jiang, W. Q., Liu, J., Zhang, J. Y., Wang, B., Frye, S., Zhang, Y., Xu, Y. H., Lei, Q. Y., Guan, K. L., Zhao, S. M., and Xiong, Y. (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of  $\alpha$ -ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30
  26. Blaschke, K., Ebata, K. T., Karimi, M. M., Zepeda-Martínez, J. A., Goyal, P., Mahapatra, S., Tam, A., Laird, D. J., Hirst, M., Rao, A., Lorincz, M. C., and Ramalho-Santos, M. (2013) Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226
  27. Chen, J., Guo, L., Zhang, L., Wu, H., Yang, J., Liu, H., Wang, X., Hu, X., Gu, T., Zhou, Z., Liu, J., Liu, J., Wu, H., Mao, S. Q., Mo, K., Li, Y., Lai, K., Qi, J., Yao, H., Pan, G., Xu, G. L., and Pei, D. (2013) Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat. Genet.* **45**, 1504–1509
  28. Minor, E. A., Court, B. L., Young, J. I., and Wang, G. (2013) Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J. Biol. Chem.* **288**, 13669–13674
  29. Cartron, P. F., Nadaradjane, A., Lepape, F., Lallier, L., Gardie, B., and Vallette, F. M. (2013) Identification of TET1 partners that control its dna-demethylating function. *Genes Cancer* **4**, 235–241
  30. Müller, U., Bauer, C., Siegl, M., Rottach, A., and Leonhardt, H. (2014) TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.* **42**, 8592–8604
  31. Deplus, R., Delatte, B., Schwinn, M. K., Defrance, M., Méndez, J., Murphy, N., Dawson, M. A., Volkmar, M., Putmans, P., Calonne, E., Shih, A. H., Levine, R. L., Bernard, O., Mercher, T., Solary, E., Uhr, M., Daniels, D. L., and Fuks, F. (2013) TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* **32**, 645–655
  32. Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., Sugawara, A., and Yokoyama, A. (2014) TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes Cells* **19**, 52–65
  33. Shi, F. T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M. A., Wan, M., and Songyang, Z. (2013) Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *J. Biol. Chem.* **288**, 20776–20784
  34. Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., and Pasini, D. (2013) Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol. Cell* **49**, 645–656
  35. Zhang, Q., Liu, X., Gao, W., Li, P., Hou, J., Li, J., and Wong, J. (2014) Differential Regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked  $\beta$ -N-acetylglucosamine transferase (OGT). *J. Biol. Chem.* **289**, 5986–5996
  36. Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013) TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493**, 561–564
  37. Hanover, J. A., Yu, S., Lubas, W. B., Shin, S. H., Ragano-Caracciola, M., Kochran, J., and Love, D. C. (2003) Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Arch. Biochem. Biophys.* **409**, 287–297
  38. Harwood, K. R., and Hanover, J. A. (2014) Nutrient-driven O-GlcNAc cycling—think globally but act locally. *J. Cell Sci.* **127**, 1857–1867
  39. Jost, K. L., Rottach, A., Mildner, M., Bertulat, B., Becker, A., Wolf, P., Sandoval, J., Petazzi, P., Huertas, D., Esteller, M., Kremmer, E., Leonhardt, H., and Cardoso, M. C. (2011) Generation and characterization of rat and mouse monoclonal antibodies specific for MeCP2 and their use in X-inactivation studies. *PLoS ONE* **6**, e26499
  40. Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellingner, K., Hasenöder, S., Wang, M., Qin, W., Söding, J., Spada, F., and Leonhardt, H. (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS ONE* **6**, e16627
  41. Liu, N., Wang, M., Deng, W., Schmidt, C. S., Qin, W., Leonhardt, H., and Spada, F. (2013) Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PLoS ONE* **8**, e62755
  42. Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikova, K., Müller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., Carell, T., and Vermeulen, M. (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159
  43. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670
  44. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
  45. Sharma, K., D'Souza, R. C., Tyanova, S., Schaab, C., Wisniewski, J. R., Cox, J., and Mann, M. (2014) Ultra-deep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.* **8**, 1583–1594
  46. Deleted in proof
  47. Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y. G. (2010) TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* **20**, 1390–1393
  48. Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y. G., Zhu, J., Wang, P., and Xu, Y. (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **155**, 1545–1555
  49. Hanover, J. A., Krause, M. W., and Love, D. C. (2012) Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. *Nat. Rev. Mol. Cell Biol.* **13**, 312–321
  50. Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A., Rappsilber, J., and Helin, K. (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348
  51. Butkinaree, C., Park, K., and Hart, G. W. (2010) O-Linked  $\beta$ -N-acetylglucosamine (O-GlcNAc): extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim. Biophys. Acta* **1800**, 96–106
  52. Myers, S. A., Panning, B., and Burlingame, A. L. (2011) Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9490–9495
  53. Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P., and Walker, S. (2011) Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564–567
  54. Jinek, M., Rehwinkel, J., Lazarus, B. D., Izaurrealde, E., Hanover, J. A., and Conti, E. (2004) The superhelical TPR-repeat domain of O-linked N-acetylglucosamine transferase exhibits structural similarities to importin  $\alpha$ . *Nat. Struct. Mol. Biol.* **11**, 1001–1007
  55. Andrés-Bergós, J., Tardío, L., Larranaga-Vera, A., Gómez, R., Herrero-Beaumont, G., and Largo, R. (2012) The increase in O-linked N-acetylglucosamine protein modification stimulates chondrogenic differentiation both *in vitro* and *in vivo*. *J. Biol. Chem.* **287**, 33615–33628
  56. Olivier-Van Stichelen, S., and Hanover, J. A. (2014) X-inactivation normalizes



## 2. Results

### ***Phosphorylation and O-GlcNAcylation of TET Proteins***

- O-GlcNAc transferase levels and generates an O-GlcNAc-depleted Barr body. *Front. Genet.* **5**, 256
57. Loenarz, C., and Schofield, C. J. (2009) Oxygenase catalyzed 5-methylcytosine hydroxylation. *Chem. Biol.* **16**, 580–583
58. Wu, H., D'Alessio, A. C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., and Zhang, Y. (2011) Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389–393
59. Sievers, F., and Higgins, D. G. (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* **1079**, 105–116
60. Rottach, A., Kremmer, E., Nowak, D., Leonhardt, H., and Cardoso, M. C. (2008) Generation and characterization of a rat monoclonal antibody specific for multiple red fluorescent proteins. *Hybridoma* **27**, 337–343

## SUPPLEMENTAL DATA

### Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the glycosyltransferase OGT<sup>\$#</sup>

Christina Bauer<sup>1</sup>, Klaus Göbel<sup>1</sup>, Nagarjuna Nagaraj<sup>2</sup>, Christian Colantuoni<sup>1</sup>, Mengxi Wang<sup>1</sup>, Udo Müller<sup>1</sup>, Elisabeth Kremmer<sup>3</sup>, Andrea Rottach<sup>1\*</sup>, Heinrich Leonhardt<sup>1,4\*</sup>

## LEGENDS

### Supplemental data S1: TET protein sequences

Annotated amino acid sequence of the murine TET constructs used for PTM mapping experiments.

Uniprot entries: Q3URK3 (TET1), Q4JK59 (TET2), Q8BG87 (TET3)

green: catalytic residues (FeII binding)

orange: Phosphorylation

pink: O-GlcNAcylation

blue: alternative splicing (variation compared to Uniprot sequence)

grey: sequences not covered by MS (no information about PTMs available)

### Supplemental data S2 (separate pdf): MS/MS spectra of O-GlcNAcylated TET peptides

### Supplemental data S3 (separate pdf): MS/MS spectra of phosphorylated TET peptides

Spectra have been annotated and exported with the MaxQuant Viewer.

### Supplemental table S4 (Excel file)

MaxQuant output tables of all detected TET peptides: unmodified (first sheet), O-GlcNAcylated (second sheet), and phosphorylated (third sheet). Summary statistics of the occupancies discussed in the main text are provided in the fourth sheet. Numbering of modified residues (column "positions within proteins") refers to Uniprot TET protein sequence and might differ from the positions used in the rest of the paper which refer to the used splicing variants cloned from murine tissues as specified in the experimental procedures. The summary statistics sheet provides both types of residue numbering .

**S1: TET PROTEIN SEQUENCES****TET1**

MSRSRPAKPSKSVKTKLQKKKDIQMKTKTSKQAVRHGASAKAVNPGKPKQ 50  
 LIKRRDGKKETEDKTPTPAPSFLTRAGAARMNRDRNQVLFQNPDSLTCNG 100  
 FTMALRRTSLSWRLSQRPVVTPKPKKVPPSKKQCTHNIQDEPGVKHSEND 150  
 SVPSQHATVSPGTENGEQNRCLVEGESQEITQSCPVFEERIEDTQSCISA 200  
 SGNLEAEISWPLEGTHCEELLSHQTSNECTSPQECAPLPQRSTSEVTSQ 250  
 KNTSNQLADLSSQVESIKLSDPSNPNTGSDHNGFPDSSFRIVPELDLKC 300  
 MPLDESUYPTALIRFILAGSQPDVFDTKPQEKTLITTPEQVGSHPNQVLD 350  
 ATSVLGQAFSTLPLQWGFSGANLVQVEALGKGSDSPEDLGAITMLNQOET 400  
 VAMMDRNPATPDLPIFLPKPPNTVATYSSPLLGPPEHSSTSCGLEVQGAT 450  
 PILTLDGHTPQLPPNPESSSVPLVIAANGTRAEKQFGTSLFPAVPQGFT 500  
 VAAENEVQHAPLDTLQGSQAAPSKLEGEISRVISITGSADV KATAM SMPVT 550  
 QASTSPPCNSTPPMVERRKRKACGVCEPCQQKANCGETYCKNRKNSHQ 600  
 ICKKRKCEVLKKKPEATSQAQVTKENKRPQREKKPKVLKTDFNNKPVNGP 650  
 KSESMDCSRRGHGEEEQRLDLITHPLENVRKNAGGMTGIEVEKWAPNKKS 700  
 HLAEGQVKGSCDANLTGVENPQPSDDKQQTNPSPFAQTIRNGMKNVHC 750  
 LPTDTHLPLNKLNHHEEFSKALGNSSKLLTDPSNCKDAMSVTSSGGECDH 800  
 LKGPRTLLFQKPGLNCRSGAEPTIFNNHPNTHSAGSRPHPEKVPNKEP 850  
 KDGSPVQPSLLSLMKDRRLTLEQVVAIEALTQLSEAPSESSSPSKPEKDE 900  
 EAHQKTASLLNSCKAILHSVRKDLQDPNVQGGKGLHHD TVVFNGQNRTFKS 950  
 PDSFATNQALIKSQGYPSPTAEKKGAAGRAPFDGFENSHPLPIESHNL 1000  
 ENCSQVLSCDQNLSSHDPSCQDAPYSQIEEDVAAQLTQLASTINHINA EV 1050  
 RNAESTPESLVAKNTKQKHSQEKRMVHQKPPSSTQTKPSVPSAKPKKAQK 1100  
 KARATPHANKRKKKPPARSSQENDQKKQEQLAIEYSKMHDIWMSSKFQRF 1150  
 GQSSPRSFPVLLRNIPVFNQILKPVTQSKTPSQHNELFPPIINQIKFTRNP 1200  
 ELAKEKVKVEPSDSLPTCQFKTESGGQTFAPADNSQGQPMVSVNQEAHP 1250  
 LPQSPPSNQCANIMAGAAQTQFHLGAQENLVHQIPPPTLPGTSPDTLLPD 1300  
 PASILRKGGKVLHFDGITVVTEKREAQTSSNGPLGPTTDSAQSEFKESIMD 1350  
 LLSKPAKNLIAGLKEQEAAPCDGCGGTQKEKGPPYYTHLGAGPSVA AVREL 1400  
 METRFGQKGKAIRIEKIVFTGKEGKSSQGCPVAKWVIRRS GPPEKLI CLV 1450  
 RERVDHHCSTAVIVVLILLWEGIPRLMADRLYKELTENLRSYSGHPTDRR 1500  
 CTLNKKRTCTCQGIDPKTCGASFSGCSWSMYFNGCKFGRSENPRKFRLA 1550  
 PNYPLHNYYKRITGMSSEGSDVKTGWIIPDRKTLISREEKQLEKNLQELA 1600  
 TVLAPLYKQMAPVAYQNQVEYEEVAGDCRLGNEEGRPFSGVTCMD FCAH 1650  
 SHKDIHNMHNGSTVVCTLIRADGRDTCPEDEQLHVLPLYRLADTDEF GS 1700  
 VEGMKAKIKSGAIQVNGPTRKRRLRFTEPVPRCGKRAKMKQNHNKSGTAG 1750  
 LRRKRISASPKGAPGSHNTKSFSSASSTSHLVKDESTDFCPLQASSAETS 1800  
 TCTYSKTASGGFAETSSILHCTMPSGAHSGANAAAGECTGTVQPAEVA AH 1850  
 PHQSLPTADSPVHAEPLTSPSEQLTSNQSNQQLPLLSNSQKLASCQVEDE 1900

RHPEADEPQHPEDDNLPQLDEFWSDSEEIYADPSFGGVAIAPIHGSVLIE 1950  
 CARKELHATTSLRSPKRGVPFRVSLVIFYQHKS LNKP NHGFDINKIKCKCK 2000  
 KVTKKKPADRECPDVSP EANLSHQIPSRVASTLTRDNVTVSPYSLTHVA 2050  
 GPYNRWV\* 2058

## TET2

MEQDRITTHAEGTRLSPFLIAPPSPISHTEPLAVKLQNGSPLAERPHPEVN 50  
 GDTKWQSSQSCYGISHMKGSSSHESPHEDRGYSRCLQNGGIKRTVLEPS 100  
 LSGLHPNKILKLDQKAKGESNIFEESQERNHGKSSRQPNVSGLSDNGEPV 150  
 TSTTQESSGADAFPTRNYNGVEIQVLNEQEGEKGRSVTLLKNKIVLMPNG 200  
 ATVSAHSEENTRGELLEKTQCYPDCVSIQVSTASHVNTTPSSQAAILSH 250  
 EIPQPSLTSAQINFSTSSQLPPEPAAMVTKACDADNASKPAIVPGTCP 300  
 FQKAHQQK S ALDIGP S RAENKTIQGSME LFAEEYYPSSDRNLQASHGSS 350  
 EQYSKQKETNGAYFRQSSKFPKD I SPTTVTPPSQSLLAPRLVLQPPLEG 400  
 KGALNDVALEEHHDPNRSNRTLLREGKIDHQPKTSSSQSLNPSVHTPNP 450  
 PLMLPEQHQNDCGSPSP EKS RKMSEYLMYYLPNHGHSGLQEH SQYLMGH 500  
 REQEIPKDANGKQIQG SVQAAPGWIELKAPNLHEALHQTKRKDI S LHSVL 550  
 HSQTGPVNQMSSKQSTGNVNMPGGFQRLPYLQKTAQPEQKAQMYQVQVNQ 600  
 GPSPGMGDQHLQFQKALYQECIPRTDPSS EAH PQAPSV PQYHFQQRVNPS 650  
 SDKHLSQQATETQRLSGFLQHTPQTQASQTPASQNSNFPQICQQQQQQQQ 700  
 QLQRKNKEQMPQTFSHLQGSNDKQREGSCFGQIKVEESFCVGNQYSKSSN 750  
 FQTHNNTQGGLEQVQNINKNFPYSKILTPNSSNLQILPSNDTHPACEREQ 800  
 ALHPVGSKTSNLQNMQYFPNNTPNQDVHRCFQEQAQK PQQASSLQGLKD 850  
 RSQGES PAPPAAEAQQRYLVHNEAKALPVPEQGSQTQTPPQKDTQKHAA 900  
 LRWLLLQKQEQQQTQSSQPGHNQMLRPIKTEPVSKPSSYRYPLSPPQENM 950  
 SSRIKQEISSPSRDNGQPKSIIETMEQHLKQFQLKSLCDYKALTLKSQKH 1000  
 VKVPTDIQAAESENHARAAEPQATKSTDCSVLDDVSESDTPGEQSQNGKC 1050  
 EGCNPDKDEAPYYTHLGAGPDVAAIRTLMEERYGEKGKAIRIEKVIYTGK 1100  
 EGKSSQGCPIAKWVYRRSSEEEKLLCLVRVRPNHTCETAVMVIAIMLWDG 1150  
 IPKLLASELYSELTDILGKCGICTNRRCSQNETKKKQSPRNCCCQGENP 1200  
 ETCGASF SFGCSWSMYNGCKFARSKKPRKFR LHGAEPKEEERLGSHLQN 1250  
 LATVIAPIYKKLAPDAYNNQVEFEHQAPDCCLGLKEGRPFSGVTACLDFS 1300  
 AHSRDQQNMPNGSTVVVTLNREDNREVGAKPEDEQFHVLPYIIAPEDE 1350  
 FGSTEGQEKKIRMGSI EVLQSFRRRRVIRIGELPKSCKKKAEPKKAATK 1400  
 AARKHSSLENCSSRTEKGKSSSHTKLMENASHMKQMTAQPOLSGPVIRQP 1450  
 PTLQRHLQQGQRPPQPPQPPQPTTPQPQPQPHIMP GNSQSVGSHCSG 1500  
 STSVYTRQPTPHSPYPSSAHTSDIYGD TNHVN FYPTSSHASGSYLNPSNY 1550  
 MNPYLGLLNQNNQYAPFPYNGSV PVDNGSPFLGSYSPQAQSRDLHRYPNQ 1600  
 DHLTNQNLPPIH TLHQQTFGDSPSKYLSYGNQNMQRDAFTNSTLKP NVH 1650  
 HLATFSPYPTPKMDSHFMGAASRSPYSHPHTDYKTSEHHLPSHTIISYTA 1700  
 AASGSSSSSHAFHNKENDNIANGLSRVLPGFNHDRTASAQELLYSLTGSQ 1750  
 EKQPEVSGQDAAAVQEIEYWS DSEHNFQDPCIGGVAIAPTHGSILIECAK 1800  
 CEVHATTKVNDPDRNHPTRI SLVLYRHKNLFLPKHCLALWEAKMAEKARK 1850  
 EEECGKNGSDHVSQKNHGKQEKREPTGPQEPSYLRFIQSLAENTGSVTTD 1900  
 STVTTSPYAFTQVTGPYNTFV\*

## 2. Results

### TET3

MDSGPVYHGDSRQLSTSGAPVNGAREPAGPGLLGAAGPWRVDQKPDWEAA 50  
SGPHTHAARLEDAHDLVAFSAVAEAVSSYGALSTRLYETFNREMSREAGSN 100  
GRGPRPESCSEGSDDLDTLQTALALARHGMKPPNCTCDGPECPDFLEWLE 150  
GKIKSMAMEGGQGRPRLPGALPPSEAGLPAPSTRPPLLSSEVPQVPPLLEG 200  
LPLSQSALSIAKEKNISLQTAIAIEALTQLSSALPQPSHSTSQASCPLPE 250  
ALSPSAPFRSPQSYLRAPSWPVVPPEEHPSFAPDSPAFPPATPRPEFSEA 300  
WGTDTPPATPRNSWPVPRPSPDPMAELEQLLGSASDYIQSVFKRPEALPT 350  
KPKVKVEAPSSSPAPVPSPISQREAPLLSSEPETHQKAQTALQQHLHHR 400  
NLFLEQAQDASFTSTEPQAPGWWAPPGSPAPRPPDKPPKEKKKKPPTPA 450  
GGPVGAEKTPPGIKTSVRKPIQIKKSRSRDMQPLFLPVRQIVLEGLKPPQA 500  
SEGQAPLPAQLSVPPPASQGAASQSCATPLTPEPSLALFAPSPSGDSLPL 550  
PTQEMRSPSPMVALQSGSTGGPLPPADDKLEELIRQFEAEFGDSFGLPGP 600  
PSVPIQEPENQSTCLPAPESPFATRSPKKIKIESSGAVTVLSTTCFHSEE 650  
GGQEATPTKAENPLTPTLSGFLESPLKYLDTPTKSLLDTPAKKAQSEFPT 700  
CDCVEQIVEKDEGPYYTHLGSQPTVASIRELMEDRYGKGAIRIEKVIY 750  
TGKEGKSSRGCPIAKWVIRRHTLEKLLCLVRHRAGHHCCNAVIVILILA 800  
WEGIPRSLGDTLYQELTDTLRKYGNPTSRRCLNDDRTCACQKDPNTCG 850  
ASFSFGCSWSMYFNGCKYARSKTPRKFRLTGDNPKKEEVLRNSFQDLATE 900  
VAPLYKRLAPQAYQNQVTNEDVAIDCRLGLKEGRPFSGVTACMDFCAHAH 950  
KDQHNLYNGCTVVCTLTKEENRCVGQIPEDEQLHVLPLYKMASTDEFGSE 1000  
ENQNAKVSSGAIQVLTAFPREVRRLEPEPAKSCRQRQLEARKAAAEKKKLQ 1050  
KEKLSTPEKIKQEAELEAGVTTDPGLSLKGGLSQQSLKPSLKVEPQNHFS 1100  
SFKYSIGNAVVESYSVLGSCRPSDPYSMSSVYSYHSRYAQPGLASVNGFHS 1150  
KYTLPSFGYYGFSSNPVFPSPQFLGPSAWGHGGSGGSFEKKPDLHALHNS 1200  
LNPAYGGAEEFAELPGQAVATDNHHPIPHHQPPAYPGPKEYLLPKVPQLHP 1250  
ASRDPSPPFAQSSSCYNRSIKQEPIDPLTQAESIPRDSAKMSRTPLPEASQ 1300  
NGGPSHLWGQYSGGSPSPKRTNSVGGNNGVFPFGESPTIVPDKLNSFGA 1350  
SCLTPSHFPESQWGLFTGEGQQSAPHAGARLRGKPWSPCKFGNGTSALTG 1400  
PSLTEKPPWGMGTGDFNPALKGGPGFQDKLWNPVKVEEGRIPPGANPLDK 1450  
AWQAFGMPLSSNEKLFGALKSEEKLWDPFSLEEGTAEPPSKGVVKEEKS 1500  
GPTVEEDEEELWSDSEHNFLDENIGGVAVAPAHCSILIECARRELHATTP 1550  
LKKPNRCHPTRISLVFYQHKNLNQPNHGLALWEAKMKQLAERARQRQEEA 1600  
ARLGLGQQEAKLYGKKRKWGGAMVAEPQHKEKKGAIPTRQALAMPTDSAV 1650  
TVSSYAYTKVTGPYSRWI\* 1669

## **2.4 A modular open platform for systematic functional studies under physiological conditions**

published May 24, 2015, in *Nucleic Acids Research*

# A modular open platform for systematic functional studies under physiological conditions

Christopher B. Mulholland<sup>1</sup>, Martha Smets<sup>1</sup>, Elisabeth Schmidtmann<sup>1</sup>, Susanne Leidescher<sup>1</sup>, Yolanda Markaki<sup>1</sup>, Mario Hofweber<sup>1</sup>, Weihua Qin<sup>1</sup>, Massimiliano Manzo<sup>1</sup>, Elisabeth Kremmer<sup>2</sup>, Katharina Thanisch<sup>1</sup>, Christina Bauer<sup>1</sup>, Pascaline Rombaut<sup>3</sup>, Franz Herzog<sup>3</sup>, Heinrich Leonhardt<sup>1,\*</sup> and Sebastian Bultmann<sup>1,\*</sup>

<sup>1</sup>Ludwig Maximilians University Munich, Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Großhaderner Strasse 2, 82152 Planegg-Martinsried, Germany, <sup>2</sup>Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Institute of Molecular Immunology, Marchioninistrasse 25, 81377 Munich, Germany and <sup>3</sup>Gene Center and Department of Biochemistry, Ludwig Maximilians University Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany

Received April 13, 2015; Revised May 13, 2015; Accepted May 14, 2015

## ABSTRACT

**Any profound comprehension of gene function requires detailed information about the subcellular localization, molecular interactions and spatio-temporal dynamics of gene products. We developed a multifunctional integrase (MIN) tag for rapid and versatile genome engineering that serves not only as a genetic entry site for the Bxb1 integrase but also as a novel epitope tag for standardized detection and precipitation. For the systematic study of epigenetic factors, including *Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Tet1*, *Tet2*, *Tet3* and *Uhrf1*, we generated MIN-tagged embryonic stem cell lines and created a toolbox of prefabricated modules that can be integrated via Bxb1-mediated recombination. We used these functional modules to study protein interactions and their spatio-temporal dynamics as well as gene expression and specific mutations during cellular differentiation and in response to external stimuli. Our genome engineering strategy provides a versatile open platform for efficient generation of multiple isogenic cell lines to study gene function under physiological conditions.**

## INTRODUCTION

In the last decades targeted gene disruption has been a widely used approach to gain first insights into gene function. However, gene disruption studies are often hampered by high functional redundancy in mammalian systems and yield little information about the subcellular localization,

interactions and spatio-temporal dynamics of gene products. In order to gain comprehensive understanding of gene function these studies need to be complemented by more complex genetic manipulations such as fluorophore knockin, specific domain deletions or introduction of point mutations. Additionally, a systematic analysis of gene function requires application of biochemical as well as imaging techniques, which usually rely on the generation of gene specific antibodies, a technically demanding and time-consuming process.

Recently, RNA guided endonucleases (RGENs) derived from the prokaryotic Type II CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated) system have emerged as promising tools for the manipulation and modification of genetic sequences (1–4).

The specificity of RGENs is mediated by small guide RNAs (gRNAs) that bind to 20 bp within the target sequence and recruit the Cas9 nuclease to introduce a double strand break. Although this two-component system has greatly facilitated the generation of gene disruptions in bacteria, plants and mammals, concerns have been raised about considerable off-target effects (5–7). Furthermore, the low frequency of homologous recombination in mammals makes insertion of exogenous components such as fluorophore tags difficult and time-consuming.

In addition to RGENs, phage-derived serine integrases have received considerable attention as novel tools for genome engineering. Recently, Bxb1 was shown to have the highest accuracy and efficiency in a screen of fifteen candidate serine integrases tested in mammalian cells (8). Serine integrases are unidirectional, site-specific recombinases that promote the conservative recombination between phage attachment sites (*attP*) and bacterial attachment sites (*attB*)

\*To whom correspondence should be addressed. Tel: +49 89 2180 74233; Fax: +49 89 2180 74236; Email: bultmann@bio.lmu.de  
Correspondence may also be addressed to Heinrich Leonhardt. Tel: +49 89 2180 74232; Fax: +49 89 2180 74236; Email: h.leonhardt@lmu.de;



(9) with much higher recombination efficiencies (up to 80%) than the commonly used bidirectional tyrosine integrases, Cre or Flp (9–12).

In this study, we aim to combine the advantages of both RGENS and unidirectional integrases into one fast, widely applicable and flexible method. We developed a novel strategy for genome engineering based on a CRISPR/Cas assisted in-frame insertion of an *attP* site, which we refer to as the multifunctional integrase (MIN) tag. At the genetic level, the MIN-tag serves as an attachment site for the serine integrase Bxb1 that can be used to introduce a broad range of prefabricated functional cassettes into the genomic locus with high specificity and efficiency. At the protein level, the MIN-tag functions as a novel epitope tag that can be detected with a highly specific monoclonal antibody and used for immunoprecipitation as well as immunofluorescence experiments. To demonstrate the versatility of the strategy, we generated MIN-tagged murine embryonic stem cell (mESC) lines for a variety of major epigenetic factors, including *Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Tet1*, *Tet2*, *Tet3* and *Uhrf1*. We created a toolbox of vectors for Bxb1-mediated recombination to generate isogenic cell lines harboring knockout cassettes, fluorescent protein fusions, enzymatic tags and specific mutations; all derived from a single entry cell line ensuring maximal biological comparability. We demonstrate the power of this strategy using proximity-dependent protein labeling to identify novel interactors of TET1 in mESCs as well as to systematically study the subcellular localization, binding kinetics and protein expression dynamics of the *de novo* methyltransferase DNMT3B during epiblast differentiation.

## MATERIALS AND METHODS

### Western blotting and immunoprecipitation

Western blot analysis was performed using the following primary antibodies: anti-DNMT1, anti-DNMT3a (Imgenex, 64B1446); anti-DNMT3b (Abcam, 52A1018); anti-UHRF1 (13); anti-TET1, anti-TET2 and anti-TET3 (14); anti-GFP antibody (Roche, 11814460001); anti- $\beta$ -Actin (Sigma, A5441); anti-SNF2H (Abcam, ab22012). Blots were probed with anti-rat (Jackson ImmunoResearch, 112-035-068), anti-mouse (Sigma, A9044) and anti-rabbit (Biorad, 170-6515) secondary antibodies conjugated to horseradish peroxidase (HRP) and visualized using an ECL detection kit (Pierce). An anti-mouse antibody conjugated to Alexa 488 (Life Technologies, A21202) was used for fluorescence detection of western blots using the Typhoon 9400 (GE Healthcare) imaging system.

For immunoprecipitation,  $\sim 1 \times 10^6$  *Dnmt1<sup>attP/attP</sup>*, *Dnmt3b<sup>attP/attP</sup>* or wt cells were harvested in ice cold phosphate buffered saline (PBS), washed twice and subsequently homogenized in 200  $\mu$ l lysis buffer (20 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM PMSF, 0.5% NP40). After centrifugation (10 min, 14 000 g, 4°C) the supernatant was adjusted with dilution buffer (20 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM PMSF) to a final volume of 300  $\mu$ l. A total of 50  $\mu$ l were mixed with sodium dodecyl sulphate (SDS)-containing sample buffer (referred to as input (I)). For pull-downs, 100  $\mu$ l (4  $\mu$ g) of either 5A10 DNMT1 antibody (15) or

the newly generated MIN-tag antibody 1E1 was added to the cell lysates and incubated 2 h at 4°C. For pull-down of immunocomplexes, 40  $\mu$ l of protein G agarose beads (GE Healthcare, Freiburg, Germany) equilibrated in dilution buffer were added and incubation continued for 2 h. After centrifugation (2 min, 5000  $\times$  g, 4°C) 50  $\mu$ l of the supernatant was collected (referred to as flow-through (FT)) while the remaining supernatant was removed. The beads were washed twice with 1 ml dilution buffer containing 300 mM NaCl. After the last washing step, the beads were resuspended in 50  $\mu$ l Laemmli buffer and boiled for 10 min at 95°C. For immunoblot analysis, 3% of the input and the flow-through as well as 30% of the bound (B) fraction were separated on a 10% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and subjected to western blot analysis.

### Immunofluorescence staining and microscopy

Immunostaining was performed as described previously (16). Briefly, cells cultured on coverslips were fixed with 4% paraformaldehyde for 10 min, washed with PBST (PBS, 0.02% Tween20) and permeabilized with PBS supplemented with 0.5% Triton X-100. Both primary and secondary antibody were diluted in blocking solution (PBST, 2% BSA, 0.5% fish skin gelatin). Coverslips with cells were incubated with primary and secondary antibody solutions in dark humid chambers for 1 h at RT; washings after primary and secondary antibodies were done with PBST. Following secondary antibody incubations, cells were post-fixed with 4% paraformaldehyde for 10 min. For DNA counterstaining, coverslips were incubated in a solution of DAPI (2  $\mu$ g/ml) in PBS. Coverslips were mounted in antifade medium (Vectashield, Vector Laboratories) and sealed with colorless nail polish.

For immunolabeling, the following primary antibodies were used: anti-DNMT1 (15); anti-DNMT3A (Imgenex, 64B1446); anti-DNMT3B (Abcam, 52A1018); anti-UHRF1 (13); anti-TET1, anti-TET2 (14); GFP-Booster-ATTO488 (Chromotek). The secondary antibodies were anti-rabbit conjugated to DyLight fluorophore 594 (Jackson ImmunoResearch, 711-505-152), anti-mouse conjugated to Alexa 488 (Life Technologies, A21202), anti-rat conjugated to Alexa 488 (Life Technologies, A21208) or Alexa 594 (Life Technologies, A21209).

Single optical sections or stacks of optical sections were collected using a Leica TCS SP5 confocal microscope equipped with Plan Apo 63 $\times$ /1.4 NA oil immersion objective and lasers with excitation lines 405, 488, 561 and 633 nm.

Live cell imaging experiments were performed on an UltraVIEW VoX spinning disc microscope assembled to an Axio Observer D1 inverted stand (Zeiss) and using a 63 $\times$ /1.4 NA Plan-Apochromat oil immersion objective. The microscope was equipped with a heated environmental chamber set to 37°C and 5% CO<sub>2</sub>. Fluorophores were excited with 488 nm or 561 nm solid-state diode laser lines. Confocal image series were typically recorded with 14-bit image depth, a frame size of 1024  $\times$  1024 pixels and a pixel size of 110 nm. z-stacks of 12  $\mu$ m with a step size of 1  $\mu$ m were recorded every 30 min for about 24 h or for the live

cell series of *Dnmt3b<sup>attP/attP</sup>* every hour for 60 h. To avoid photodamage of the cells, the AOTF of the laser was set to low transmission values of 6–10%. Binning was set to 2×.

### Super-resolution microscopy

Super-resolution images were obtained with a DeltaVision OMX V3 3D-SIM microscope (Applied Precision Imaging, GE Healthcare), equipped with a 60×/1.42 NA PlanApo oil objective and sCMOS cameras (Olympus). A z-step size of 125 nm was used during acquisition. SI raw data were reconstructed and deconvolved with the SoftWorX 4.0 software package (Applied Precision). FIJI and Photoshop CS5.1 (Adobe) were used for image processing and assembly.

### Antigen preparation, immunization, generation of hybridomas and ELISA screening

For the translated attP peptide, the MIN antigen (attP peptide) was designed with the following sequence SGQPPRSQWCTVQT-Cys. Peptides were synthesized, HPLC purified and coupled to OVA (Peps4LifeSciences-Anette Jacob; Heidelberg). Lou/c rats were immunized subcutaneously and intraperitoneally with a mixture of 50 µg peptide-OVA, 5 nmol CPG oligonucleotide (Tib Molbiol, Berlin), 500 µl PBS and 500 µl incomplete Freund's adjuvant. A boost without adjuvant was given 6 weeks after primary injection. Fusion of the myeloma cell line P3 × 63-Ag8.653 with the rat immune spleen cells was performed using polyethylene glycol 1500 (PEG 1500, Roche, Mannheim, Germany). After fusion, the cells were plated in 96 well plates using RPMI1640 with 20% fetal calf serum, penicillin/streptomycin, pyruvate, non-essential amino acids (Gibco) supplemented by hypoxanthine-aminopterin-thymidine, (HAT) (Sigma, St Louis, MO, USA). Hybridoma supernatants were tested in a solid-phase immunoassay. Microliter plates were coated with avidin (3 µg/ml, Sigma) over night. After blocking with 2% FCS in PBS, plates were incubated with biotinylated MIN peptide at a concentration of 0.2 µg/ml in blocking buffer. After washing the plates, the hybridoma supernatants were added. Bound rat mAbs were detected with a cocktail of HRP-labeled mouse mAbs against the rat IgG heavy chains, thus avoiding IgM mAbs (α-IgG1, α-IgG2a, α-IgG2b (ATCC, Manassas, VA, USA), α-IgG2c (Ascension, Munich, Germany). HRP substrate conversion was visualized with ready to use TMB (1-Step™ Ultra TMB-ELISA, Thermo). MIN-tag clone 1E1 (rat IgG1) was stably subcloned and further characterized.

A set of 25 rat derived hybridoma supernatants were tested for specificity against an integrated *attP* peptide in the *Dnmt1* locus using both western blot analysis and high content microscopy. Western blots were prepared as mentioned previously. Each supernatant was used in a 1:10 dilution. Blots were probed with an anti-rat secondary antibody conjugated to HRP.

Cells were prepared for immunofluorescence as described above, with the exception that cells were fixed on a 96-well Cell Carrier® plate (Greiner). Cells in individual wells were incubated with the various hybridoma supernatants (1:100)

for 1 h. As a secondary antibody, anti-rat conjugated to Alexa 488 (Life Technologies, A21208) was used. Nuclei were counterstained using DAPI. Images of stained cells were acquired automatically with an Operetta high-content imaging system using a 40× air objective (PerkinElmer). DAPI and ATTO488 coupled antibodies were excited and their emissions recorded using standard filter sets. Exposure times were 10 and 400 ms for DAPI and ATTO488, respectively. All monoclonal antibodies described in this study are available upon request.

The MIN antibody are available via [http://human.bio.lmu.de/\\_webtools/MINtool/AB.info.html](http://human.bio.lmu.de/_webtools/MINtool/AB.info.html).

### DNA methylation analysis

For the analysis of DNA methylation levels, genomic DNA was isolated using the QIAamp DNA Mini Kit (QIAGEN). Bisulfite treatment was performed using the EZ DNA Methylation-Gold™ Kit (Zymo Research Corporation) according to the manufacturer's protocol. Subsequently, the major satellite repeats sequence was amplified using the primers described in (17). The biotinylated polymerase chain reaction (PCR) products of the second PCR were analyzed by pyrosequencing (Varionostic GmbH, Ulm, Germany).

### Targeting donor and plasmid construction

Plasmid sequences can be found in Supplementary Table S6. Targeting donor constructs were either synthesized as ssDNA oligonucleotides (Integrated DNA Technologies) or produced by amplifying 300 to 200 bp long homology arms with the respective external and internal primer sets (Supplementary Table S2). These PCR products of the 5' and 3' homology arms were pooled and an overlap extension PCR with the external primers was performed to yield the final targeting fragments. The gRNA vector was synthesized at Eurofins MWG Operon based on the sequences described (3). The subcloning of targeting sequences was performed by circular amplification. The surrogate reporter (pSR) was generated by inserting *in vitro* annealed DNA oligos via AsiSI and NruI into pCAG-mCh (18). eGFP was amplified using the primers eGFP-F and eGFP-R and sequentially cloned into pCAG-mCh-NruI linker to generate the pSR construct. Reporters were generated by subcloning *in vitro* annealed DNA oligos containing CRISPR target sites into KpnI and NheI digested pSR. The attB-GFP-knockin construct was generated from R6K-NFLAP (19) by ligation free cloning (20) rearranging the backbone sequences into the artificial intron and introducing the attB site 5' of the GFP open reading frame (ORF), removing its start codon. The attB-GFP-Poly(A) and attB-mCh-Poly(A) constructs were created by amplifying the GFP ORF including the stop codon and SV40 Poly(A) signal from pCAG-eGFP-IB and inserted into the attB-LAP-tag backbone by ligation free cloning. The attB-mCh-Poly(A)-mPGK-PuroR construct was generated by subcloning the mPGK-PuroR sequence from pPthc-Oct3/4 (21) and ligating it into the EcoRV site of the attB-mCh-Poly(A) construct. The attB-GFP-Poly(A)-mPGK-NeoR was produced by first exchanging

the PuroR in pPthc-Oct3/4 with NeoR from pEGFP-C1 (22) using HindIII. The combined mPGK-NeoR was then subcloned into the attB-GFP-Poly(A) vector via the same EcoRV site mentioned previously. The attB-GFP-Dnmt1cDNA-Poly(A), attB-GFP-Tet1cDNA-Poly(A) and attB-GFP-Dnmt3b1cDNA-Poly(A) constructs were generated by inserting the appropriate cDNAs from constructs reported previously (17,23–24) via AsiSI/NotI sites into the attB-GFP-Poly(A) and attB-mCh-Poly(A) vectors respectively. The attB-GFP-Dnmt3b6-Poly(A), attB-GFP-Tet1-d1–1363-Poly(A), attB-GFP-Tet1-d833–1053-Poly(A), attB-GFP-Tet1-d833–1363-Poly(A) vectors were produced via circular amplification with overlap extension primers using the above mentioned attB-GFP-Dnmt1/Dnmt3b1/Tet1cDNA-Poly(A) constructs as templates.

The attB-GFP-Dnmt3b6-Poly(A)-mPGK-NeoR and attB-mCh-Dnmt3b1-Poly(A)-mPGK-PuroR integration constructs were created by inserting the Dnmt3b6 and Dnmt3b1 sequences (from attB-GFP-Dnmt3b6-Poly(A) and attB-GFP-Dnmt3b1-Poly(A)) using AsiSI/NotI sites into attB-GFP-Poly(A)-mPGK-NeoR and attB-mCh-Poly(A)-mPGK-PuroR vectors, respectively.

All constructs described in this study are available via Addgene or via [http://human.bio.lmu.de/\\_webtools/MINTool/](http://human.bio.lmu.de/_webtools/MINTool/).

### Cell culture

J1 ESCs were maintained on gelatin-coated dishes in Dulbecco's modified Eagle's medium supplemented with 16% fetal bovine serum (FBS, Biochrom), 0.1 mM  $\beta$ -mercaptoethanol (Invitrogen), 2 mM L-glutamine, 1  $\times$  MEM Non-essential amino acids, 100 U/ml penicillin, 100  $\mu$ g/ml streptomycin (PAA Laboratories GmbH), 1000 U/ml recombinant mouse LIF (Millipore) and 2i (1  $\mu$ M PD032591 and 3  $\mu$ M CHIR99021 (Axon Medchem, Netherlands), referred to as ESC medium. Differentiation of naive pluripotent stem cells to epiblast-like cells was performed according to the protocol of (25). Briefly, J1 ESCs were maintained in the ground state in Geltrex (Life Technologies) coated flasks and cultured in N2B27 (50% neurobasal medium (Life Technologies), 50% DMEM/F12 (Life Technologies), 2 mM L-glutamine (Life Technologies), 0.1 mM  $\beta$ -mercaptoethanol, N2 supplement (Life Technologies), B27 serum-free supplement (Life Technologies) containing 2i and 1000 U/ml LIF 100 U/ml Penicillin-streptomycin) for at least three passages before differentiation. To differentiate naive ESCs into epiblast-like cells, cells were replated in N2B27 differentiation medium containing 10 ng/ml Fgf2 (R&D), 20 ng/ml Activin A (R6D) and 0.1  $\times$  Knockout Serum Replacement (KSR)(Life Technologies). Time point 0 h in differentiation time-course experiments corresponds to the time N2B27 differentiation medium was added to cells.

### Generation of MIN-tagged and Bxb1-mediated knockin cell lines

To produce MIN-tagged cell lines,  $5 \times 10^5$  cells were dissociated and seeded in 0.2% gelatin (Sigma-Aldrich) coated p35 plates. After 3 h, cells were transfected with 2  $\mu$ g of

the MIN-tag donor/homology ssDNA oligo or PCR product, 0.5  $\mu$ g gRNA construct, 0.5  $\mu$ g surrogate reporter construct and 1  $\mu$ g Cas9 using Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions. For Bxb1-mediated recombination of attB constructs,  $5 \times 10^5$  cells were transfected with 1  $\mu$ g pCAG-NLS-HA-Bxb1 expression plasmid ((26) addgene 51271), 1  $\mu$ g of the respective attB construct and 0.5  $\mu$ g Bxb1 surrogate reporter. For both MIN-Tagging and Bxb1-mediated recombination, cells were dissociated, resuspended in ESC medium 48 h post transfection and then analyzed and sorted with a FACS Aria II (Becton Dickinson). For MIN-tagging, enrichment of cells with RGEN activity was accomplished by single-cell sorting GFP and mCherry positive cells into 96-well plates (Falcon) containing 150  $\mu$ l of ESC medium. For Bxb1-mediated recombination, cells with Bxb1 activity were enriched for by single-cell sorting GFP positive cells into 96-well plates. Alternatively for Bxb1-mediated integration using antibiotic selection, cells were replated into p150 plates with ESC medium containing G418 (0.5 mg/ml, AppliChem) and puromycin (1  $\mu$ g/ml, AppliChem) 48 h post transfection.

### Identification of MIN-tagged and Bxb1-mediated knockin cell lines with restriction fragment analysis and PCR screening

After ~7 days (until colonies were readily visible), plates from single-cell sortings were screened for colony growth. Surviving colonies were dissociated and individually replated onto two 96-well plates. Genomic DNA was isolated from one plate after 2–3 days, while the second plate remained in culture. To identify MIN-tagged clones, the region surrounding the ATG (or stop codon in the case of C-terminal tagging) was PCR amplified using the appropriate external and screening primers (Supplementary Table S2). For restriction fragment analysis, 10  $\mu$ l of these PCR products were digested with either HincII or SacII and then analyzed on 1.5% agarose gels. PCRs of positive clones were confirmed by Sanger sequencing. To screen for Bxb1-mediated recombination, we employed a three-primer PCR strategy using the respective external primers flanking the MIN-tagged locus and an attL-specific primer (Supplementary Figure S3A, Table S2). For Bxb1-mediated integrations using antibiotic selection, mESC colonies were picked, dissociated using trypsin and plated into individual wells on 96-well plates ~7 days after starting antibiotic selection. Genomic DNA isolation and screening PCRs were performed as described above. Clones harboring the desired MIN-tag insertion or Bxb1-mediated integration were expanded, frozen and stored in liquid nitrogen.

All cell lines are available at [http://human.bio.lmu.de/\\_webtools/MINTool/cell\\_lines.html](http://human.bio.lmu.de/_webtools/MINTool/cell_lines.html).

### Genomic DNA isolation for PCR

Cells were lysed in multi-well plates by the addition of 50  $\mu$ l lysis buffer (10mM Tris/HCl pH 7.4, 10mM EDTA, 10mM NaCl, 50 $\mu$ g/ml Proteinase K, 1.7  $\mu$ M SDS) per well. The Plates were subsequently incubated at  $-80^\circ\text{C}$  for 15 min, followed by 3 h at  $56^\circ\text{C}$ . Heat inactivation of Proteinase K

was performed by incubation at 85°C for 20 min. The resulting crude DNA lysates were directly subjected to PCR.

### BioID

BioID experiments were performed after (27) using extracted crude nuclei (adapted from (28)) as input material. In brief, cells were cultured for 48 h with or without addition of 50  $\mu$ M biotin. Cell pellets ( $\sim 4 \times 10^7$  cells) were washed once in buffer A (10 mM HEPES/KOH pH 7.9, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>) and resuspended in buffer A containing 0.15% NP-40 and 1 $\times$  protease inhibitor (SERVA). Samples were homogenized using a pellet pestle. After centrifugation, crude nuclei pellets were washed once with PBS. Crude nuclei were resuspended in BioID-lysis buffer (0.2% SDS, 50 mM Tris/HCl pH 7.4, 500 mM NaCl, 1 mM DTT, 1 $\times$  protease inhibitor), supplemented with 2% Triton X-100 and subjected to sonication twice using a Branson Sonifier 450 (15% amplitude, 0.3 s pulse, 0.6 s pause, total time 30 s). Samples were diluted 1:1 with 50 mM Tris/HCl pH 7.4 after the first sonication step. Pulldown of biotinylated proteins was performed overnight at 4°C with rotation using M-280 Streptavidin Dynabeads (Life Technologies) for subsequent mass spectrometry or Streptactin-Superflow agarose beads (IBA) for SDS-PAGE analysis, respectively. Beads were washed with wash buffer 1 (2% SDS), wash buffer 2 (0.1% desoxycholic acid, 1% Triton X-100, 1 mM EDTA, 500 mM NaCl, 50 mM HEPES/KOH pH 7.5) and wash buffer 3 (0.5% desoxycholic acid, 0.5% NP-40, 1 mM EDTA, 500 mM NaCl, 10 mM Tris/HCl pH 7.4) followed by two washing steps with 50 mM Tris/HCl pH 7.4. For SDS-PAGE analysis, proteins were silverstained after (29).

### Digest of proteins and sample preparation for LC-MS/MS

On-beads digest of proteins was performed as described in (28). All steps were carried out at room temperature. Beads were resuspended in 2 M Urea in Tris/HCl pH 7.5, reduced with 10 mM DTT for 20 min and subsequently alkylated with 50 mM chloroacetamide for 20 min. A total of 0.25  $\mu$ g Pierce Trypsin Protease (Thermo Scientific) was added for 2 h. Beads were collected by centrifugation and the resulting peptide supernatant was further incubated overnight with addition of 0.1  $\mu$ g trypsin. Peptides were desalted using StageTips (30).

### LC-MS/MS and data analysis

Peptides were reconstituted in 20  $\mu$ l mobile phase A (2% v/v acetonitrile, 0.1% v/v formic acid) and analyzed by tandem mass spectrometry using a EASY-nLC 1000 nano-HPLC system connected to a LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). About 2–4  $\mu$ l of the peptide mixture were separated onto a PepMap RSLC column (75  $\mu$ m ID, 150 mm length, C18 stationary phase with 2  $\mu$ m particle size and 100 Å pore size, Thermo Fisher Scientific) and introduced into the mass spectrometer at a flow rate of 300 nl/min running a gradient from 5 to 35% mobile phase B (98% v/v acetonitrile, 0.1% v/v formic acid). Ion source and transmission parameters of the mass spectrometer were set to spray voltage = 2 kV, capillary temperature = 275°C. The mass spectrometer was operated in

data-dependent mode, selecting up to 10 precursors from a MS1 scan (resolution = 60 000) in the range of m/z 250–1800 for collision-induced dissociation (CID). Singly (+1) charged precursor ions and precursors of unknown charge states were rejected. CID was performed for 10 ms using 35% normalized collision energy and the activation q of 0.25. Dynamic exclusion was activated with a repeat count of one, exclusion duration of 30 s, list size of 500 and the mass window of  $\pm 10$  ppm. Ion target values were 1 000 000 (or maximum 10 ms fill time) for full scans and 10 000 (or maximum 100 ms fill time) for MS/MS scans, respectively. Raw data were analyzed using MaxQuant Version 1.5.2.8 (31) using the MaxLFQ label free quantification algorithm (32) and the match-between-runs functionality. UniprotKB MOUSE.fasta was used as a reference database (33). A maximum of two missed cleavages and a false discovery rate of 1% were set as parameters. Oxidation of methionine and biotinylation were searched as variable modifications and carbamidomethylation of cysteine residues as fixed modification. For statistical analysis, the Perseus software version 1.5.1.6 was used (31). Significance was tested using a two sided Student's *t*-test and a permutation based FDR calculation. GO enrichment analysis was performed with the *Gene Ontology enrichment analysis and visualization tool* (GORilla, (34)). A *P*-value < 0.01 was considered significant.

### FRAP

Live cell imaging and FRAP experiments were typically performed on an UltraVIEW VoX spinning disc microscope with integrated FRAP PhotoKinesis accessory (PerkinElmer) assembled to an Axio Observer D1 inverted stand (Zeiss) and using a 63 $\times$ /1.4 NA Plan-Apochromat oil immersion objective. The microscope was equipped with a heated environmental chamber set to 37°C. Fluorophores were excited with 488 nm (exposure time: 400 ms, laser power: 15%) or 561 nm (exposure time: 450 ms, laser power: 30%) solid-state diode laser lines. Confocal image series were typically recorded with 14-bit image depth, a frame size of 256  $\times$  256 pixels and a pixel size of 110 nm. For photobleaching experiments, the bleach regions, typically with a diameter of 2  $\mu$ m, were manually chosen to cover the chromocenters. Photobleaching was performed using one iteration with the acousto-optical tunable filter (AOTF) of the 488 nm laser line set to 100% transmission. Typically, 10 pre-bleach images were acquired at a rate of 1 s per timepoint and 60 post-bleach frames were recorded at a rate of 10 s per timepoint. Data correction, normalization and quantitative evaluations were performed by automated processing with ImageJ (<http://rsb.info.nih.gov/ij/>) using a set of newly developed macros followed by calculations in Excel.

## RESULTS

### A fast and efficient strategy to generate MIN-tagged genomic loci

Our novel genome engineering strategy relies on the CRISPR/Cas-assisted insertion of the MIN-tag sequence into the open reading frame of a target gene either directly



downstream of the start codon or upstream of the stop codon (Figure 1A and Supplementary Figure S2H). Neither regulatory regions nor gene structure are altered, leading to preservation of the endogenous expression pattern and post-transcriptional processing of the gene of interest.

Since epigenetic processes undergo dramatic changes during early embryonic development and are tightly regulated, we tested the efficacy and versatility of our method by targeting the DNA modifying enzymes *Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Tet1*, *Tet2* and *Tet3* as well as the chromatin binding protein *Uhrf1* in mESCs (Figure 1D). We generated targeting donors containing the 48 bp MIN-tag sequence flanked by short homology arms (200–300 bp for PCR-based donors or 76 bp for single stranded DNA oligos). We next designed specific gRNAs to target sequences located either in close proximity to or overlapping the start or stop codon of the respective genes. As scarless integration of the MIN-tag requires a resistance free selection strategy we used a surrogate reporter assay to enrich for cells that express an active Cas9:gRNA complex by fluorescence-activated cell sorting (FACS) (Figure 1B and C). In this reporter assay, the target sequence is inserted between the ORF of mCherry (mCh) and GFP thereby disrupting the reading frame of the fusion. GFP is expressed only when the target sequence is cleaved by a specific and active Cas9:gRNA complex, which causes small, frameshifting insertions or deletions by non-homologous end joining (NHEJ) restoring the reading frame of the fluorescent protein (35). For each targeting, we co-transfected mESCs with a mixture of surrogate reporter construct, gRNA vector, Cas9 expression plasmid and the specific targeting MIN-tag donor fragment. After single cell sorting of GFP positive cells and expansion of the resulting colonies, we isolated genomic DNA by a fast and simplified in-well lysis protocol to screen for positive clones by PCR and analytical restriction digest. This allows the identification of hetero- and homozygous insertions already at this stage (Supplementary Figure S1D). Combined, all targeting yielded positive clones with an average efficiency of 3% for homozygous and 1% for heterozygous insertions (Supplementary Table S1). All targeted genes were expressed normally and subcellular localization as well as enzymatic activity was not disrupted in comparison to wild-type (wt) cells (Supplementary Figures S1 and S2). In addition, the possibility of C-terminal tagging (see *Uhrf1* (C); Figure 1D and Supplementary Figure S2H) allows the MIN-tag to be used in cases where N-terminal targeting disturbs protein function.

Taken together, these results demonstrate that the MIN-tag can efficiently be integrated at precise genomic locations using a CRISPR/Cas assisted, fluorescence based selection strategy.

#### Generation of a highly specific monoclonal antibody recognizing the MIN epitope

Insertion of the MIN-tag into the ORF of target genes leads to expression of a small peptide that does not occur in the mammalian proteome (Figure 2A). This unique feature allowed us to generate a highly specific monoclonal antibody against MIN-tagged proteins. Immunofluorescence (IF) stainings of a mixed *Dnmt1*<sup>attP/attP</sup> and wt culture dis-

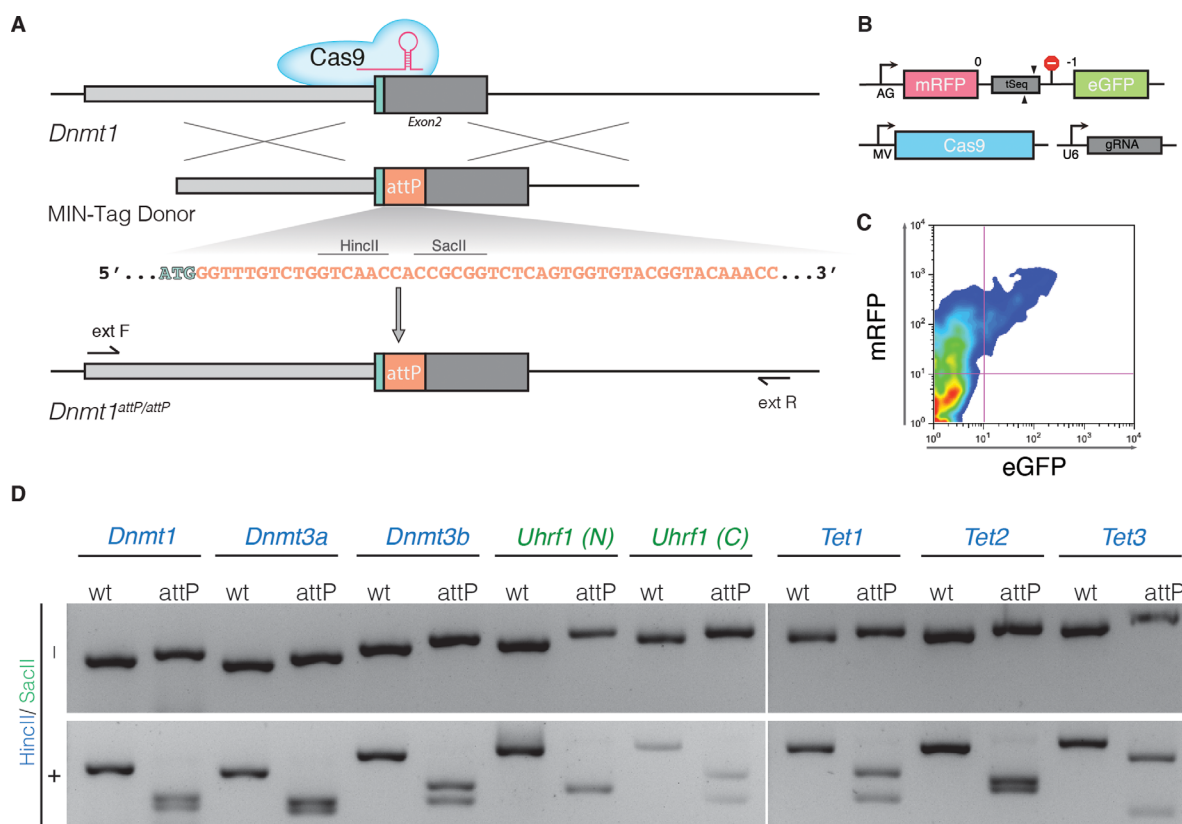
tinguished single MIN-tagged cells and colonies from wt cells, demonstrating the high specificity of the anti-MIN antibody (Figure 2B). Pull-down experiments in *Dnmt1*<sup>attP/attP</sup> cell extracts showed a quantitative enrichment of DNMT1 in the bound fraction (Figure 2C). Furthermore, pull-down of DNMT3B using the anti-MIN antibody efficiently coprecipitated SNF2H, a known interactor of DNMT3B, in protein extracts of *Dnmt3b*<sup>attP/attP</sup> cells, but not in wt control extracts (Figure 2D) (36).

Collectively, these data show that the MIN-tag can be utilized as a universal epitope tag for IF and immunoprecipitation (IP), thus allowing the investigation of localization and molecular interactions of MIN-tagged proteins.

#### Functionalization of MIN-tagged genes by Bxb1-mediated recombination

To demonstrate the versatility of the MIN-tag as a Bxb1 integration site, we constructed a toolbox of functional cassettes, which we recombined into the MIN-tagged locus of the maintenance DNA methyltransferase *Dnmt1* (*Dnmt1*<sup>attP/attP</sup>). First, we generated a knockout vector carrying the *attB* site directly in front of the ORF of GFP followed by a stop codon and a polyadenylation signal (*attB*-GFP-Poly(A), Figure 3A) that we transfected together with a codon-optimized Bxb1 expression construct in the *Dnmt1*<sup>attP/attP</sup> cell line. Successful recombination events were identified by GFP expression and single cells sorted by FACS (Figure 3B). We designed a multiplex PCR strategy that takes advantage of the unique *attL* site generated by successful recombination to facilitate identification of positive clones and their zygosity (Figure 3D and Supplementary Figure S3A). PCR screening of sorted clones revealed that the *attB*-GFP-Poly(A) construct had been successfully integrated into both alleles in 13 (56.5%) clones (Supplementary Table S3). Of those, we examined three clonal cell lines all of which exhibited no residual expression of DNMT1 by western blot analysis and IF (Figure 3F; Supplementary Figure S3B and C). For functional characterization, we analyzed DNA methylation levels at major satellite repeats, one of the main substrates for DNA methylation activity of DNMT1 during replication (37,38). Due to the loss of the maintenance DNA methyltransferase in the *Dnmt1*<sup>KO/KO</sup> clones, a severe hypomethylation was observed at this sequence (Figure 3E). Taken together, our *attB*-GFP-Poly(A) vector proved to be a valuable tool to generate genetically-defined gene knockouts in MIN-tagged cell lines.

Second, we designed a GFP knockin construct that can be used to generate in-frame GFP fusions of MIN-tagged genes. To avoid disruption of the gene locus and preserve the endogenous splicing sites, we placed the bacterial backbone sequences into an artificial intron splitting the GFP ORF into two exons (19) (Figure 3A). After recombination and FACS sorting for GFP expressing cells, the GFP knockin construct integrated in both alleles of the *Dnmt1* locus in 13 clones (41.9%), without altering physiological DNMT1 expression levels (Figure 3G, Supplementary Figure S3D and Table S3). Live cell imaging of *Dnmt1*<sup>GFP/GFP</sup> cells revealed a normal localization of GFP-DNMT1 throughout the cell cycle (15,24) (Supplementary Figure S3E), demonstrating



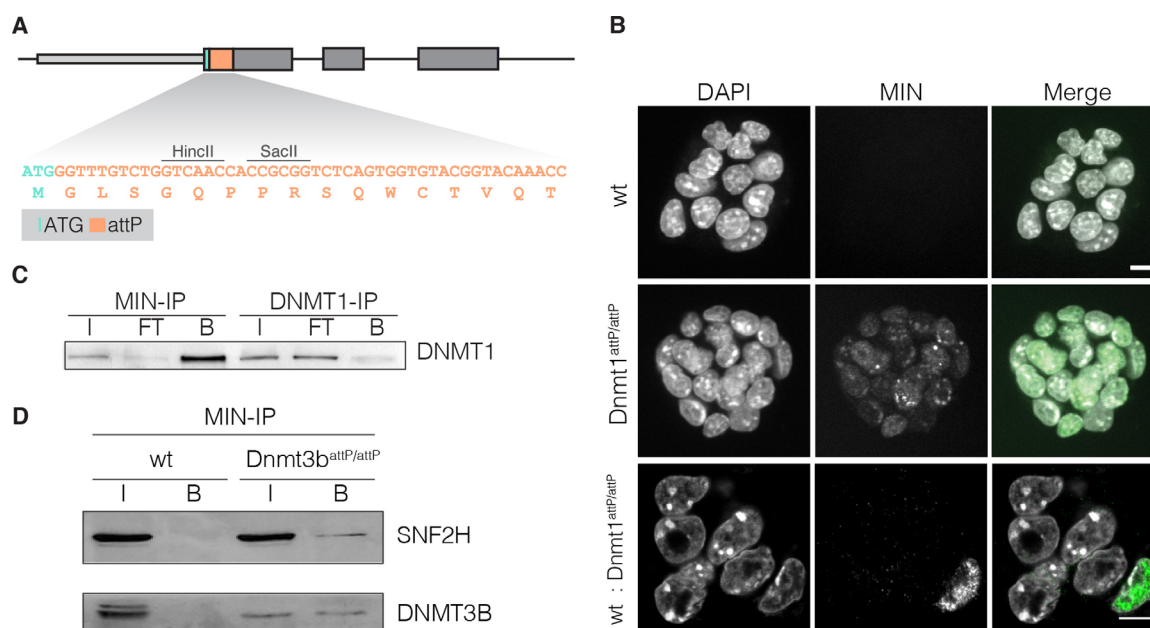
**Figure 1.** Generation of MIN-tagged cell lines. (A) Schematic overview of MIN-tag insertion into the *Dnmt1* locus via CRISPR/Cas assisted gene editing. The MIN-tag donor harbors the *attP* site and homology to the genomic sequence 5' and 3' of the start codon. Integration is facilitated by double strand breaks created by Cas9 directed to the target sequence by a specific gRNA. Restriction enzyme recognition sites used for screening in this study are indicated above the *attP* sequence. (B) Schematic overview of the surrogate reporter used to enrich for cells expressing a functional Cas9 complex. The respective Cas9 target sequence (tSeq) is placed downstream of mRFP followed by a stop codon and an out-of-frame GFP ORF. This surrogate reporter is transfected into the cells together with a vector expressing Cas9 and a U6 driven gRNA expression cassette. (C) Cells that express a functional Cas9 complex can then be identified by expression of GFP and enriched via FACS. (D) Screening PCRs followed by restriction digest with HincII or SacII of all generated MIN-tagged cell lines. (N) and (C) refer to N- and C-terminal tagging, respectively.

that DNMT1 regulation was not impaired. Albeit only at low frequencies, Bxb1 has been shown to damage recombination sites (8). Therefore, we sought to confirm that the Bxb1-mediated recombination of the GFP cassette at the MIN-tagged locus occurred without error via site-specific recombination. We sequenced the region flanking the *attL* site in the *Dnmt1*<sup>GFP/GFP</sup> cell line (Supplementary Figure S4) and determined that the GFP cassette was accurately integrated in a scarless fashion. In summary, this attB-GFP vector is suited to express GFP fusion proteins from the endogenous promoter preserving physiological regulation and splicing of the target gene.

Finally, we investigated whether the MIN-tag can be used to generate cell lines expressing mutants of the target gene for functional screenings or disease modeling. We cloned the cDNA of *Dnmt1* into the *attB-GFP-Poly(A)* construct in-frame with GFP and performed recombination as described above. We identified 10 (66.6%) clones in which integration had occurred, of which 9 (60%) were homozygous for the *Dnmt1* cDNA knockin (Supplementary Table S3).

Expression analysis by western blot and live cell imaging revealed that the endogenous DNMT1 protein was completely replaced by the *Dnmt1* mini gene product and exhibited normal localization (Figure 3H, Supplementary Figure S3F).

All in all, we show that MIN-tagged entry cell lines can be efficiently functionalized with a flexible toolbox of attB-vectors to generate gene knockouts, N-terminal fusion constructs such as GFP and cDNA knockins. In total, we generated 15 derivatives of our MIN-tagged cell lines so far. The efficiency of Bxb1-mediated recombination ranged from 33 to 67%, with an average of 50% (Supplementary Table S3, Figure S5). This demonstrates the efficacy of our system as well as the simplicity with which MIN-tagged cell lines can be modified and functionalized by prefabricated cassettes. The error-prone step of CRISPR/Cas-mediated insertion of the MIN-tag is necessary only once to generate an entry cell line, which can then be specifically manipulated with a variety of recombination vectors, allowing maximum biological comparability.



**Figure 2.** Application of the anti-MIN monoclonal antibody. (A) DNA sequence of the *attP* site and corresponding translated MIN peptide sequence (orange). (B) Fluorescence micrographs of wt mESCs, Dnmt1attP/attP cells and of a mixed culture (1:10) of wt and Dnmt1attP/attP cells stained with the anti-MIN antibody. DAPI is used as DNA counterstain. Scale bars represent 5  $\mu$ m. (C) IP experiments performed with anti-MIN and anti-DNMT1 antibody in Dnmt1attP/attP cell extracts (input (I), flow through (FT), bound (B)). (D) Co-IP of DNMT3B in wt and Dnmt3battP/attP cells using the anti-MIN antibody. DNMT3B co-precipitated SNF2H in Dnmt3battP/attP cells as determined by western blot.

### Using the MIN-tag strategy to study endogenous protein regulation

As elucidating the function of uncharacterized protein domains requires systematic analysis, we generated a series of deletion constructs covering the N-terminus of TET1, which we aimed to recombine into our *Tet1<sup>attP/attP</sup>* cell line (Figure 4A). However, we were unable to identify positive recombination events by FACS due to low expression of this target gene. To circumvent this problem, we developed a surrogate reporter system for Bxb1 mediated recombination that can be used to enrich for positive recombination events (Figure 3C). The Bxb1 surrogate reporter construct consists of a constitutive promoter followed by an attP site and a Poly(A) sequence. Upon transfection, Bxb1 mediates the recombination of a fluorophore (e.g. GFP) containing *attB* plasmid with the Bxb1 surrogate reporter, which results in the expression of GFP. This allows enrichment of positive recombination events, even when the MIN-tagged gene is not expressed or only at low levels.

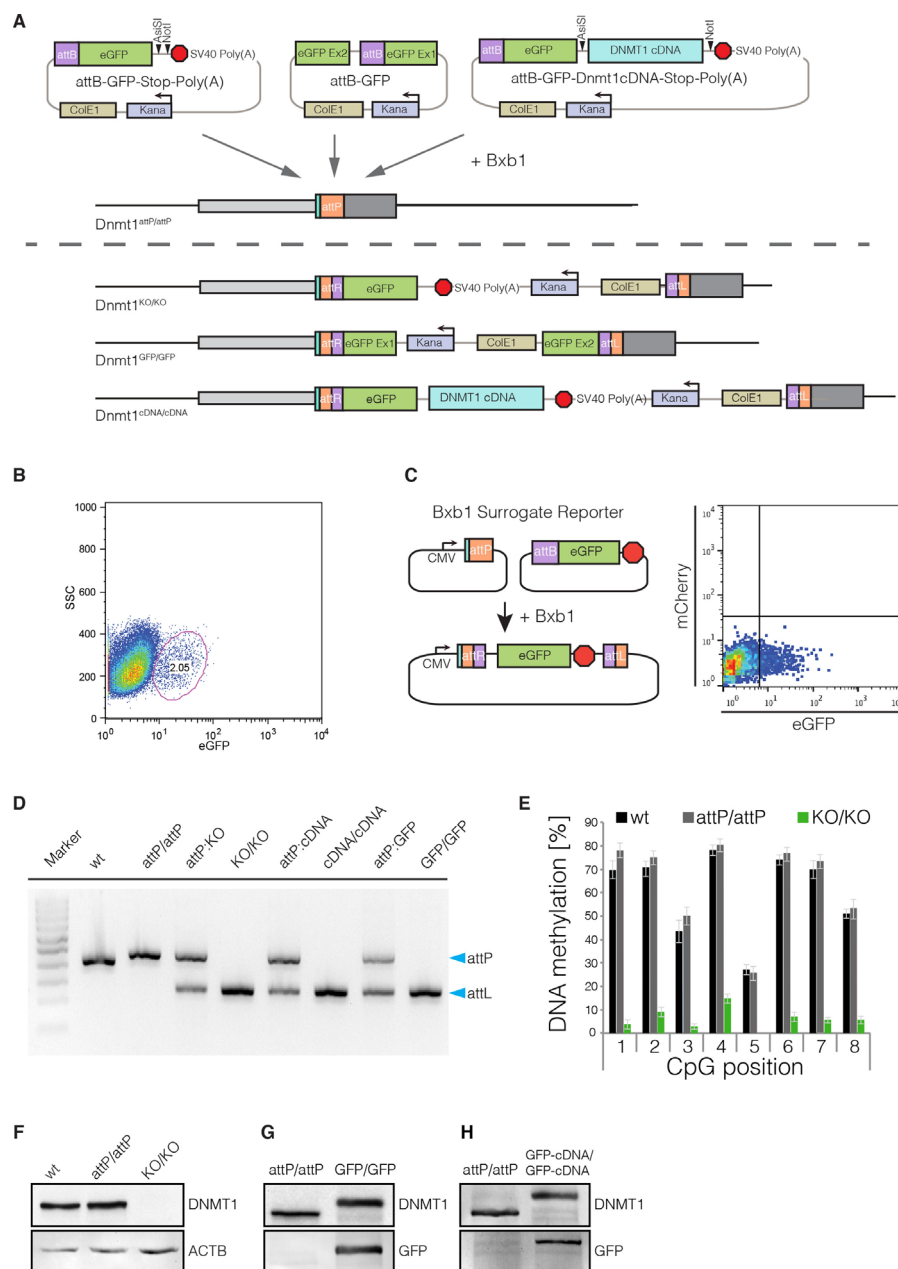
Using the Bxb1 surrogate reporter for enrichment and the above described PCR strategy for screening, we were able to generate four *Tet1* knockin cell lines expressing N-terminal deletion constructs from the endogenous promoter. Western blot analysis revealed complete replacement of wt TET1 expression by the knockin constructs (Figure 4B). These cell lines can be used for future systematic studies of the regulatory function of the TET1 N-terminus that is largely unknown so far.

Taking advantage of the MIN-tag strategy to express fusion constructs at endogenous levels, we expanded our toolbox to include a BirA\* cassette which we knocked into the *Tet1* locus (Supplementary Figure S5G). In contrast to classical IP approaches, proximity-dependent protein labeling by the promiscuous biotin ligase, BirA\* (BioID) (27), allows the characterization of the full microenvironment of a protein of interest independent of physical protein-protein interactions. This technique enabled us to pull down proteins within close proximity ( $\sim 10$  nm radius, (39)) of TET1 that were subsequently identified by LC-MS/MS (Figure 4C). We found nine proteins to be significantly enriched (40) upon addition of exogenous biotin to the culture medium of our *Tet1<sup>BirA\*/BirA\*</sup>* mESC line, including SIN3A, a known interactor of TET1 (41) (Figure 4D and E). Interestingly, these proteins are associated with chromatin modification and organization (Figure 4F). This marks the first time that the BioID method has been used in mESCs and in a non-overexpression context with the BirA\* ligase fused to the endogenous protein.

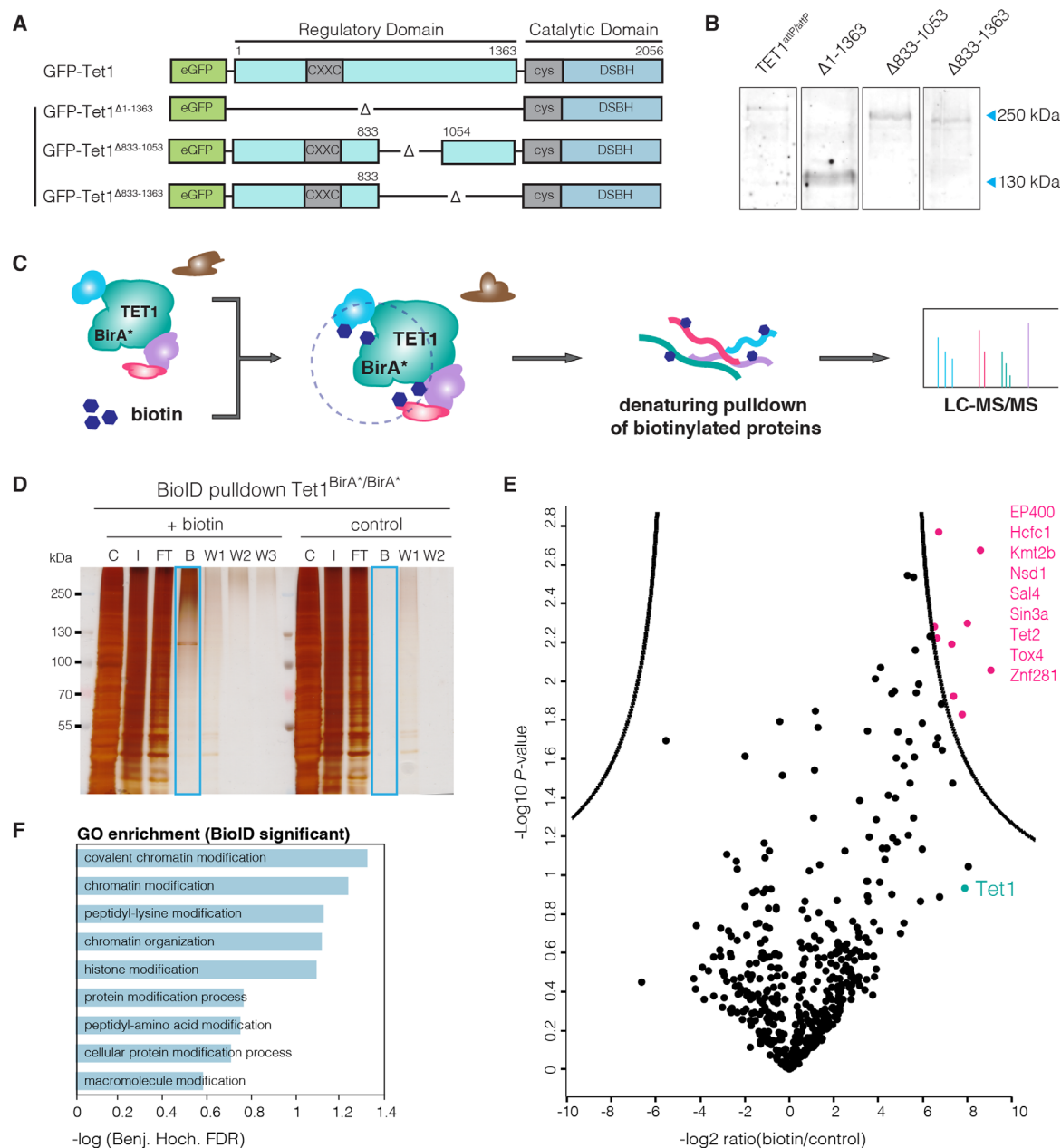
### Using the MIN-tag strategy to study dynamic cellular processes

During early embryonic development, the epigenome undergoes massive rearrangements that are precisely regulated. Knockout of the major epigenetic factors is often embryonic lethal (38,42) and over-expression studies frequently fail to reflect the tight regulation of these proteins. Therefore, more flexible and delicate genetic manipulations





**Figure 3.** Bxb1-mediated insertion of functional cassettes into the *Dnmt1* locus. (A) Schematic outline of the strategy and vectors used to create knockout, GFP knockin and cDNA knockin functionalizations of the *Dnmt1*<sup>attP/attP</sup> cell line. cDNAs can be cloned into the attB-GFP-Stop-Poly(A) vector using the 8-cutters AsiSI and NotI. (B) FACS plot depicting the gating and sorting of mESCs to enrich for cells positive for integration of the knockout cassette (2.05% of parent population) based on GFP expression. (C) The Bxb1 surrogate reporter consists of a constitutive CMV promoter followed by an attP site. If Bxb1 and attB donor plasmid containing GFP is present in the cell, recombination of the donor into the reporter leads to expression of GFP. The Bxb1 surrogate reporter can be used to enrich for successful recombination events by FACS. (D) Gel electrophoresis of the multiplex PCR for wt, *Dnmt1*<sup>attP/attP</sup> (attP/attP), *Dnmt1*<sup>KO/KO</sup> (KO/KO), *Dnmt1*<sup>cDNA/cDNA</sup> (cDNA/cDNA) and *Dnmt1*<sup>GFP/GFP</sup> (GFP/GFP) as well as 1:1 mixtures with *Dnmt1*<sup>attP/attP</sup> genomic DNA, to control for amplification biases. Blue arrows indicate expected sizes of the non-recombined (attP) and recombined allele (attL). (E) DNA methylation levels at the major satellite repeats of *Dnmt1*<sup>KO/KO</sup> cells compared to wt and *Dnmt1*<sup>attP/attP</sup> cells. (F) Western blot analysis of DNMT1 expression levels in wt, *Dnmt1*<sup>attP/attP</sup> and *Dnmt1*<sup>KO/KO</sup> cells generated by Bxb1-mediated insertion of a knockout cassette. (G) Western blot analysis of DNMT1 and GFP expression in *Dnmt1*<sup>attP/attP</sup> and homozygous GFP-knockin cells (*Dnmt1*<sup>GFP/GFP</sup>) generated by Bxb1-mediated insertion. (H) Western blot analysis of DNMT1 and GFP expression in *Dnmt1*<sup>attP/attP</sup> and *Dnmt1*<sup>cDNA/cDNA</sup> cells expressing a GFP-Dnmt1 minigene from the endogenous promoter.



**Figure 4.** Study of TET1 regulation. (A) Schematic representation of the *Tet1* cDNA constructs used for Bxb1-mediated recombination into *Tet1<sup>attP/attP</sup>* cells. (B) Western blot analysis of TET1 expression in *Tet1<sup>attP/attP</sup>* cell line and its derivatives expressing GFP-TET1 $\Delta$ 1-1363 ( $\Delta$ 1-1363), GFP-TET1 $\Delta$ 833-1053 ( $\Delta$ 833-1053) and GFP-TET1 $\Delta$ 833-1363 ( $\Delta$ 833-1363). Note that fusion to GFP increases the MW of TET1 constructs by 29 kDa. (C) Schematic representation of the BioID approach as described by Roux *et al.* (27). (D) SDS-PAGE analysis of a BioID pulldown experiment using the *Tet1<sup>BirA\*/BirA\*</sup>* cell line. Cells were cultured either without (control) or with 50  $\mu$ M biotin (+biotin). C: Cytoplasm, I: Crude nuclei input, FT: Flowthrough, B: Bound, W1-W3: Wash. (E) Volcano plot of proteins identified in the streptavidin pulldown of the TET1-BioID experiment, quantified with the MaxQuant Label-Free Quantification algorithm (32). The x-axis reflects the difference in protein abundance in the BioID pull-down compared to the negative control while the y-axis shows the logarithmized *P*-value of a student's *t*-test. Significantly enriched proteins are highlighted in pink (FDR = 0.01, S0 = 3, indicated by black line (40)). Experiments were performed in duplicates. (F) GO term enrichment of proteins identified as significant in BioID.

are needed to study the function of epigenetic factors *in vivo*. Here, we focus on the *de novo* DNA methyltransferase 3B (DNMT3B), one of the key factors during epiblast differentiation. While it has been shown that DNMT3B, in concert with DNMT3A and DNMT3L, is responsible for the global wave of *de novo* DNA methylation occurring during epiblast differentiation (42–44), little is known about its localization and protein kinetics during this developmental time period.

To address this question in a systematic fashion, we generated a homozygous GFP knockin cell line (*Dnmt3b<sup>GFP/GFP</sup>*) from the *Dnmt3b<sup>attP/attP</sup>* cell line by Bxb1-mediated recombination (Figure 5A and 6A). This allowed us to follow expression of DNMT3B under native regulatory conditions and to monitor its localization during the two-day transition from naive pluripotent ESCs to Epiblast-like cells (EpiLCs, (25)) using live cell imaging with high temporal resolution (1 image per hour).

At the naive pluripotent state, we observed very low expression levels of DNMT3B. Upon addition of differentiation medium, protein expression was strongly and uniformly upregulated reaching its maximum at 48–52 h (Figure 5B, Supplementary video 1). Overall, these findings were consistent with *Dnmt3b* mRNA levels in wt and *Dnmt3b<sup>attP/attP</sup>* cells (Figure 5C). Interestingly, we observed a highly dynamic subnuclear distribution of DNMT3B during differentiation that can be classified into three patterns (Figure 5B). (i) In the first 14 h of differentiation, DNMT3B is expressed at low levels and no clear enrichment is visible. (ii) Between 14–40 h after initiation of differentiation, DNMT3B expression is upregulated and accumulates at constitutive heterochromatin of chromocenters (CCs). (iii) After 40 h of differentiation, DNMT3B is highly expressed and localization to CCs is diminished. The above-described patterns were not related to specific cell cycle stages, indicating a differentiation stage dependent localization of DNMT3B (Supplementary Figure S6A).

To investigate the specific chromatin distribution of DNMT3B during differentiation in more detail, we performed super-resolution 3D structured illumination microscopy (3D-SIM) with the anti-MIN antibody for protein visualization. DAPI and trimethylated lysine 4 of histone 3 (H3K4me3) were used as markers of heterochromatin and euchromatin (45), respectively. In agreement with the live cell imaging experiments, DNMT3B localizes at CCs, clusters of subcentromeric regions, at the 30 h time point and shows a broader distribution at 60 h after differentiation (Figure 5D). Interestingly, the higher resolution of 3D-SIM revealed an accumulation of the signal in facultative heterochromatin at perinuclear and perinucleolar regions at the 60 h time point (Figure 5D; right panel).

DNMT3B has been shown to be responsible for the methylation of major satellite DNA, a main constituent of CCs (42,46–47). As DNMT3B is enriched at CCs between 14–40 h of differentiation, we investigated whether DNMT3B is actively methylating these sequences during this period. Therefore, we performed fluorescence recovery after photobleaching (FRAP) of GFP-DNMT3B localized at CCs. Using our *Dnmt3b<sup>GFP/GFP</sup>* cell line, we performed FRAP experiments at 35 h of differentiation. Using circular regions of interest (ROIs) that encompassed individual CCs, we monitored signal recovery for 10 min after

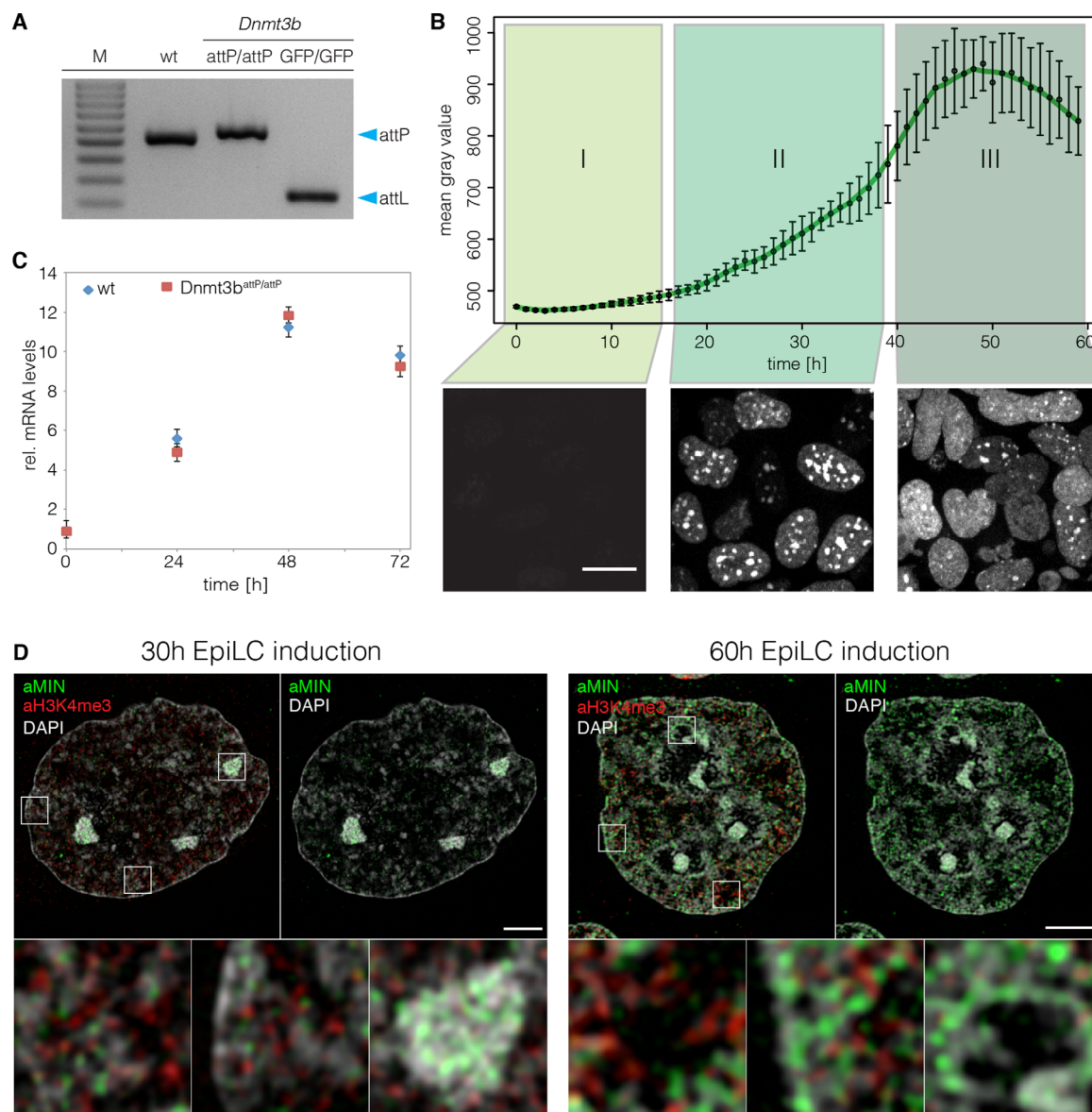
bleaching. We found that the signal exhibited a slow recovery rate ( $t_{1/2} = 42$  s) and did not recover completely. As DNA methylation has been shown to have a slow turnover rate (48,49), this suggested the immobile fraction (~20%) of DNMT3B could be catalytically active at CCs (Figure 6B and D, Supplementary Table S4). To test this hypothesis, we performed FRAP experiments on cells treated with the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine (5-azadC), which irreversibly traps DNMTs at their site of action (50). We found that 5-azadC treated CCs exhibited a large immobile fraction (~80%) suggesting that DNMT3B is actively methylating CCs at this time point. However, we were surprised to find that ~20% of DNMT3B enzyme still remained mobile (Figure 6C). Considering the long 5-azadC treatment time of 12 h this suggested that a fraction of the enzyme never engaged in catalytic reactions. As our GFP cassette preserves endogenous splicing patterns, the GFP-DNMT3B fusions used in this study represent a mixture of different protein isoforms. This prompted us to investigate the contribution of *Dnmt3b* splicing isoforms to the observed FRAP kinetics.

For *Dnmt3b*, nine splicing isoforms, all originating from the same translational start site, have been described (51). Besides the catalytically active isoform DNMT3B1, DNMT3B6 has been shown to be highly expressed in ESCs. This isoform is produced by alternative splicing, skipping exons 23 and 24, resulting in a protein that lacks several highly conserved motifs within the catalytic domain and has therefore been suggested to be inactive (52).

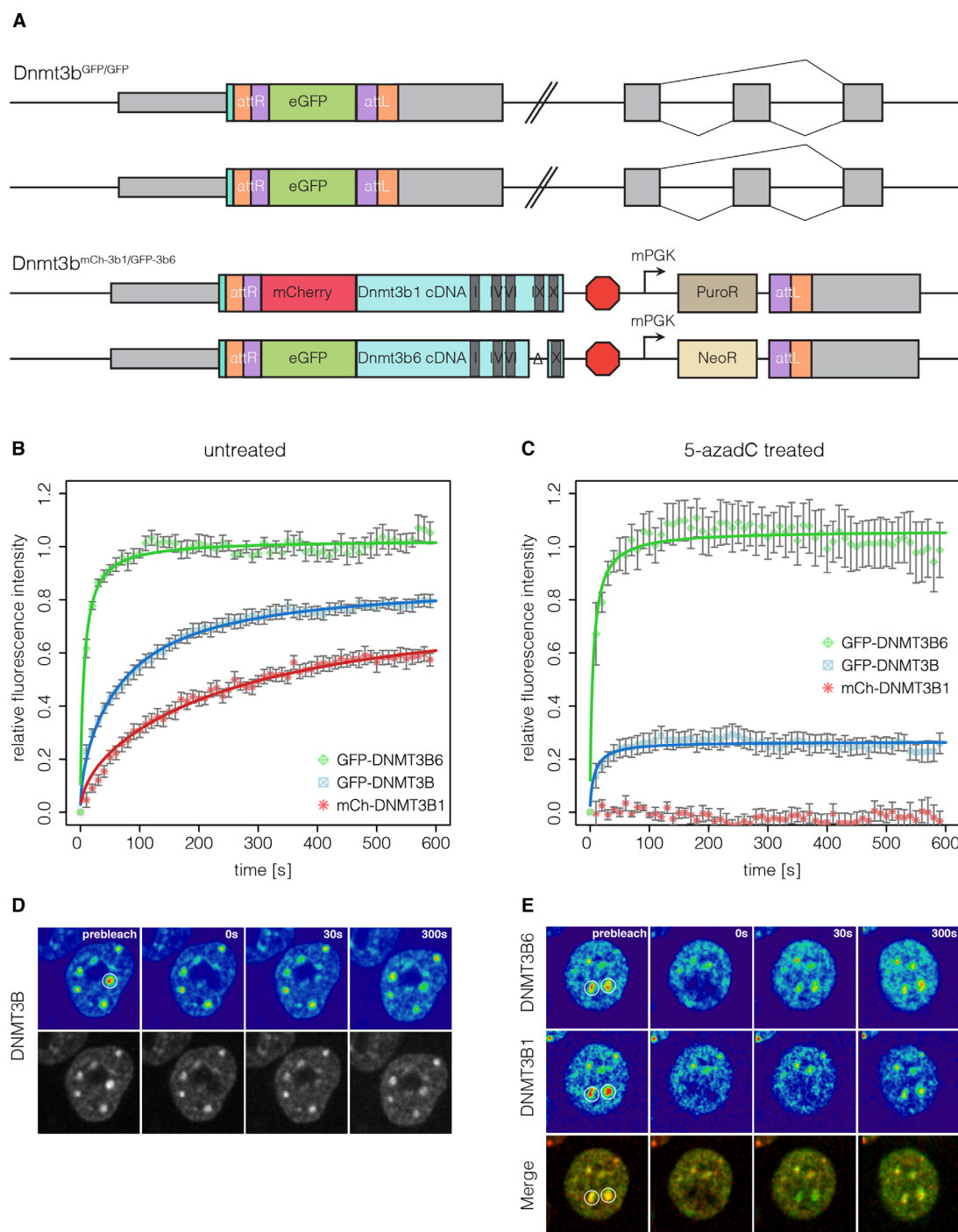
To dissect the contributions of DNMT3B1 and DNMT3B6 to the observed FRAP kinetics of *Dnmt3b<sup>GFP/GFP</sup>* cells, we generated a cell line expressing fluorescent fusions of each isoform. For this, we produced cDNA knockin constructs in which DNMT3B1 was fused to a red fluorescent protein mCherry (mCh) and DNMT3B6 was fused to GFP. To facilitate the generation of knockin cell lines expressing each isoform from one allele we equipped the *Dnmt3b1* and *Dnmt3b6* constructs with a Neomycin and Puromycin resistance cassette, respectively. We successfully established a cell line that simultaneously expressed mCh-DNMT3B1 and GFP-DNMT3B6, both under the control of the endogenous *Dnmt3b* promoter (Figure 6A, Supplementary Figure S6B), allowing us to directly compare the FRAP kinetics of DNMT3B1 and DNMT3B6 within the same cell. In the absence of 5-azadC, GFP-DNMT3B6 exhibited a fast ( $t_{1/2} = 5$  s) and complete recovery while mCh-DNMT3B1 recovered slower ( $t_{1/2} = 95$  s) (Figure 6B, Supplementary Table S4).

Intriguingly, FRAP kinetics of DNMT3B6 were not influenced by the presence of 5-azadC, supporting that it is catalytically inactive. In contrast, DNMT3B1 was completely immobilized by addition of 5-azadC exhibiting virtually no recovery after photobleaching (Figure 6C and E).

Taken together, our MIN-tag strategy enabled us to show that DNMT3B exhibits a dynamic localization to distinct chromatin regions during epiblast differentiation. Super-resolution micrographs of cells stained with anti-MIN antibodies at different time points of epiblast differentiation hint towards progression of *de novo* DNA methylation in a hierarchical fashion starting at constitutive (CCs) and progressing towards facultative (perinuclear/perinucleolar)



**Figure 5.** Spatio-temporal dynamics of DNMT3B during epiblast differentiation. **(A)** Gel electrophoresis of the multiplex screening PCR for wt, *Dnmt3b<sup>attP/attP</sup>* and *Dnmt3b<sup>GFP/GFP</sup>*. Blue arrows indicate expected sizes of the non-recombined (attP) and recombined allele (attL). **(B)** Evaluation of GFP signals during live cell imaging of *Dnmt3b<sup>GFP/GFP</sup>* cells. The graph depicts mean gray values of nuclear GFP signals. Error bars represent standard deviations ( $n > 81$ ). Lower panels show Z-projections of *Dnmt3b<sup>GFP/GFP</sup>* cells representative of the indicated time frame. Scale bar represents 10  $\mu$ m. **(C)** Quantitative real-time PCR of *Dnmt3b* mRNA levels in wt and *Dnmt3b<sup>attP/attP</sup>* cells during epiblast differentiation. **(D)** 3D-SIM nuclear mid-sections of anti-MIN (green) and anti-H3K4me3 (red) antibody distributions 30 and 60 h after induction of EpiLC differentiation combined with DAPI counterstaining (gray) in *Dnmt3b<sup>attP/attP</sup>* cells. Lower panels represent 7 $\times$  magnifications of selected boxed regions. Scale bars represent 3  $\mu$ m and 500 nm in insets.



**Figure 6.** Protein dynamics of DNMT3B and its isoforms during epiblast differentiation. (A) Schematic representation of the *Dnmt3b* genomic loci in the *Dnmt3b*<sup>GFP/GFP</sup> and the *Dnmt3b*<sup>mCh-3b1/GFP-3b6</sup> cell lines. (B) Quantitative evaluation of FRAP experiments (average of 11–14 cells) comparing GFP-DNMT3B with GFP-DNMT3B6 and mCh-DNMT3B1 in *Dnmt3b*<sup>GFP/GFP</sup> and the *Dnmt3b*<sup>mCh-3b1/GFP-3b6</sup> cell lines differentiated for 35 h. Error bars represent standard error of the mean. (C) Quantitative evaluation of FRAP experiments (average of 10–12 cells) as in (B) with cells treated with 5-azadC 12 h before imaging. (D and E) Representative images of FRAP experiments performed in (B) and (C), respectively. White circles indicate the bleach ROI with a diameter of 2  $\mu$ m.



heterochromatin. Finally, FRAP experiments revealed that the two isoforms DNMT3B1 and DNMT3B6 exhibit dramatically different DNA binding kinetics.

## DISCUSSION

Recent advances in genome engineering technology, based on TALEN and CRISPR/Cas systems, have greatly facilitated the process of manipulating genetic information. Platforms have been established that allow genome-wide gene disruption screenings for factors involved in any biological process (20,53–54). While these methods provide valuable information about the genes and pathways involved, in-depth analysis of target genes is needed to understand their function. This, in turn, requires the implementation of various genetic, cell biological and biochemical techniques. To gain meaningful insights into gene function, these techniques have to be applied under physiological conditions requiring extensive and complex genetic manipulations. Although modern genome engineering tools have made such manipulations possible, a more efficient and universal approach would be highly desirable to implement the above-mentioned techniques in a systematic manner.

The MIN-tag strategy offers a new means of rapid, efficient, yet flexible genetic manipulation of target loci. We show that CRISPR/Cas assisted insertion of the MIN-tag can be performed efficiently with short homology donors. Several studies have shown that CRISPR/Cas mediated gene targeting is associated with a significant risk of off-target cleavage, which can result in indel (insertions or deletions) formation due to NHEJ (5–7,55–56). The MIN-tag strategy requires a single nuclease assisted gene editing event, thereby keeping the likelihood of off-target effects at a minimum. Further modifications are then performed using Bxb1-mediated recombination. In contrast to the  $\phi$ C31 integrase, Bxb1 has been shown to be highly specific with virtually no unwanted genomic insertions at pseudo *attP* sites (8–9,57–58). Once a MIN-tagged cell line is established, in-frame fusion of the MIN-tag to the target gene also results in the expression of a novel epitope tag. We show that this epitope tag can be detected by a highly specific antibody, which can be used to screen for positive clones, perform co-immunoprecipitation (co-IP) experiments, as well as conventional and super resolution microscopy.

Using Bxb1 and the MIN-tag toolbox, a MIN-tagged entry cell line can be used to generate multiple isogenic derivatives within 2–3 weeks (Figure 7), without the risk of introducing off-target effects. Our collection of vectors for Bxb1 mediated recombination currently contains over 80 different plasmids (Supplementary Table S5). These prefabricated functional cassettes constitute an expandable toolbox for the simple and flexible genetic alteration of any tagged loci, without the need of locus-specific homology.

Using our stop cassette, we show that the MIN-tag strategy can be used to reliably achieve genetically defined gene disruption of MIN-tagged genes. Harboring a Poly(A) signal, insertion of this cassette efficiently eliminates target gene expression with the added advantage of precluding unwanted downstream initiation. As fluorescent protein reporters are commonly used to study spatio-temporal dy-

namics and protein kinetics in living cells, we generated a GFP knockin construct (*attB*-GFP) for Bxb1-mediated integration. GFP knockin cell lines made with this construct retain not only their endogenous expression levels but also their endogenous splicing pattern. Similarly, a BirA\* cassette can be introduced at any MIN-tagged locus to allow for proximity-dependent labeling of the microenvironment of a given protein.

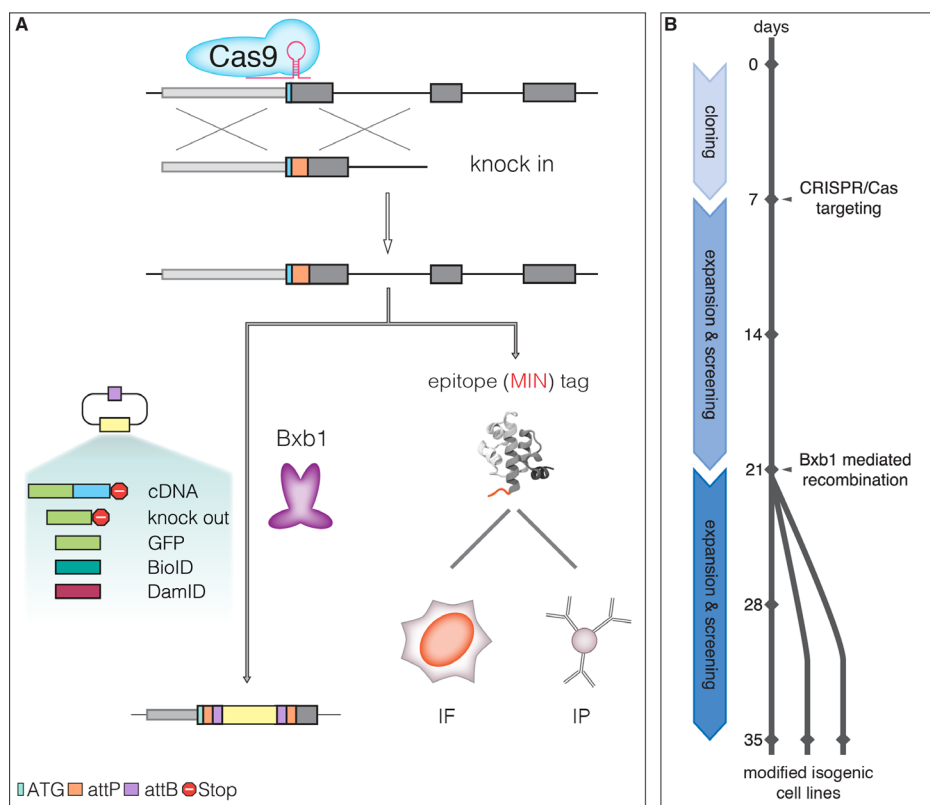
Understanding protein function often necessitates the systematic alteration of individual domains through mutations as well as deletions. Equipped with a fluorescent protein and strategic cloning sites, our cDNA knockin cassette is especially tailored for simple and expedient insertion of user-defined cDNAs. PCR-based approaches can be used to easily alter the coding sequence and quickly produce a library of gene specific cDNA mutants. These can then be inserted into target loci by Bxb1-mediated recombination, completely replacing expression of the wt gene while retaining endogenous control. While this strategy does not directly introduce the mutations into the gene locus, it offers a means of inserting and testing multiple mutant constructs in a short time frame without the need to design and perform additional nuclease-assisted targetings. This feature can be used to gain insights into the functional implications of the rapidly growing number of mutations found in cancer and disease. Likewise, the generation of large deletion mutants is easily accomplished facilitating the investigation of protein domain function and interaction mapping. This eliminates the need for excising large genomic regions or cloning long site-specific homology donors.

Obviously, the above mentioned plasmids by no means represent the extent of all possible functional cassettes. For example, MIN-tag toolbox modules allowing inducible protein stabilization or localization (59,60) as well as enzymatic labeling of DNA binding sites (DamID (61)) would greatly assist the elucidation of protein function and protein-chromatin interactions, respectively.

Employing our strategy in mESCs, we inserted the MIN-tag into the genes coding for all mammalian DNA modifying enzymes and a cofactor (*Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Tet1*, *Tet2*, *Tet3* and *Uhrf1*). These MIN-tagged cell lines as well as their functional derivatives (Supplementary Table S3) constitute a valuable resource to investigate the role of these proteins during fundamental processes such as pluripotency, cellular reprogramming, embryonic development and disease.

One gold standard method to study protein–protein interactions is co-IP. However, chromatin- or membrane-bound proteins are often barely soluble and consequently difficult to investigate by this approach. Making use of our BirA\* cassette, we investigated factors in the microenvironment of TET1, a dioxygenase that oxidizes DNA at methylated cytosines (62). Besides the known interactor SIN3A, we identify eight other proteins in proximity to TET1 that are involved in chromatin modification and organization, including the closely related TET2. This is in accordance with the findings by Costa *et al.* (63) that TET1 and TET2 have partially overlapping target sites. In conclusion, integration of the BirA\* cassette into the endogenous locus is a perfectly suited method to study dynamic protein–protein interactions.





**Figure 7.** The MIN-tag strategy. **(A)** Schematic outline of the genome engineering strategy. Small homology donors are used to insert serine integrase (*attP*) sites in-frame after the ATG codon of target genes via CRISPR/Cas assisted HR. The *attP* site is translated as a novel epitope tag suitable for IF and IP with the specific monoclonal antibody. The *attP* site is also recognized by the serine integrase Bxb1 and used for specific and directional integration of *attB*-carrying functional cassettes into the tagged gene locus. All derivatives are subjected to their endogenous gene regulation ensuring that subsequent studies are performed at physiological expression levels. **(B)** Timeline for generation of MIN-tagged genes and subsequent modification by Bxb1-mediated recombination. MIN-tagged cell lines can be generated within 2–3 weeks. These cell lines can then be modified within another 2–3 weeks to generate multiple isogenic cell lines with different functional modifications.

We also applied the MIN-tag strategy to study the *de novo* DNA methyltransferase DNMT3B during the transition from naive pluripotent ESCs to primed EpiLCs, a period of dramatic epigenetic change. While distinct patterns have been described for ESCs and somatic cells (46,64), the subnuclear distribution of DNMT3B during differentiation remains largely unknown. We discovered that DNMT3B exhibits a highly dynamic subnuclear distribution during epiblast differentiation. Our observations suggest that the global wave of *de novo* DNA methylation during epiblast differentiation follows a distinct spatio-temporal order, initiating at constitutive pericentromeric heterochromatin followed by transition to facultative heterochromatin.

Exploiting the unique possibilities of our MIN-tag strategy, we furthermore generated a cell line simultaneously expressing differentially tagged splicing isoforms of DNMT3B from different alleles. This approach revealed that the major catalytically active isoform DNMT3B1 was completely immobilized at chromocenters after 5-azadC treatment, while the FRAP kinetics of DNMT3B6 were not affected. This, to our knowledge, is the first time that FRAP

has been performed on different isoforms of a protein at endogenous expression levels in the same cell.

While this study was performed using mouse ESCs, our strategy can be applied to any cell type as long as no Bxb1 *attP* site is present in the respective genomes. The human genome is free of this entry site and introduction of the MIN-tag into cell lines such as human induced pluripotent stem cells should greatly facilitate the generation of clinically relevant disease models. Moreover, MIN-tagged mESCs could be used in blastocyst injections to generate MIN-tagged mice. Different tissues and cells could not only be used for Bxb1-mediated genetic manipulation *in vitro*, free of the limitation posed by inefficient endogenous homologous recombination, but also to study tissue specific protein regulation with the MIN-tag antibody. Furthermore, widely used cell biological model systems such as HeLa and U2OS cells as well as model organisms such as *Caenorhabditis elegans* or *Drosophila* could benefit from the versatility and efficiency of our approach.

In summary, with our combined genome engineering approach, a plethora of functional derivatives can be gener-

ated from one entry line with high efficiency. To simplify the distribution of MIN-tagged cell lines and the MIN-tag toolbox as well as to assist with the design of targeting strategies, we have developed a web-tool that is accessible at [http://human.bio.lmu.de/\\_webtools/MINtool/](http://human.bio.lmu.de/_webtools/MINtool/). As entry lines can be shared and the genetic toolbox easily expanded with new functional modules, the MIN-tag strategy represents a dynamic flexible open platform and facilitates systematic functional studies with direct biological comparability.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Ina Poser and Tony Hyman (Max Planck Institute of Molecular Biology & Genetics, Dresden) for providing R6K-NFLAP construct, George Church (Harvard Medical School, Boston) for providing the Cas9 expression construct, Kerry Tucker (Ruprecht-Karls-University, Heidelberg) for providing wt ESCs and Pawel Pelczar (University of Zürich) for providing pCAG-NLS-HA-Bxb1. Furthermore, we thank Robert Engelmeier and Michael Soutschek for help with cell line generation, IF and antibody characterization. We also thank Andy Spiegl, Gregor Jessberger and Jan Langkabel for help with website development. CBM gratefully acknowledges the Life Science Munich Graduate School. ES, CB and KT gratefully acknowledge the International Max Planck Research School for Molecular and Cellular Life Sciences.

## FUNDING

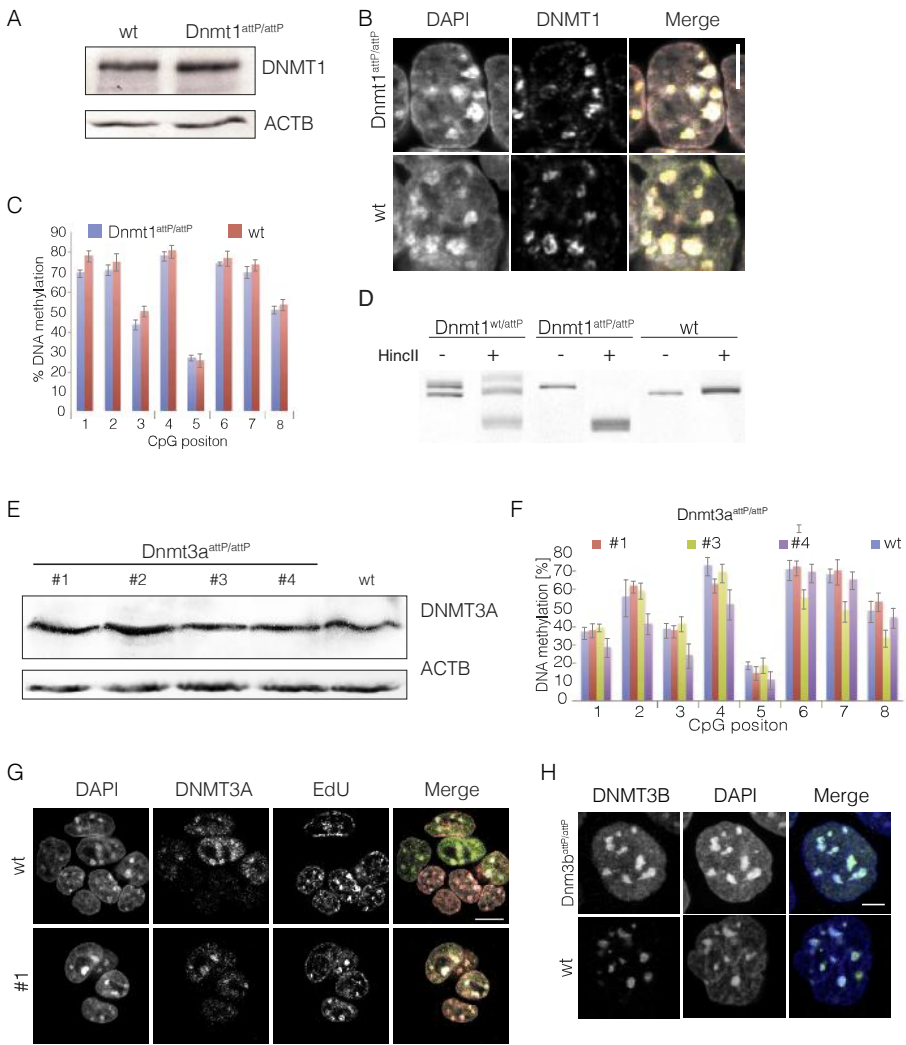
Deutsche Forschungsgemeinschaft [SFB 1064 to H.L., E.K. and GRK 1721 to H.L., F.H.]; Funding for open access charge: Deutsche Forschungsgemeinschaft [Collaborative Research Center SFB 1064/A17; GRK 1721].  
*Conflict of interest statement.* None declared.

## REFERENCES

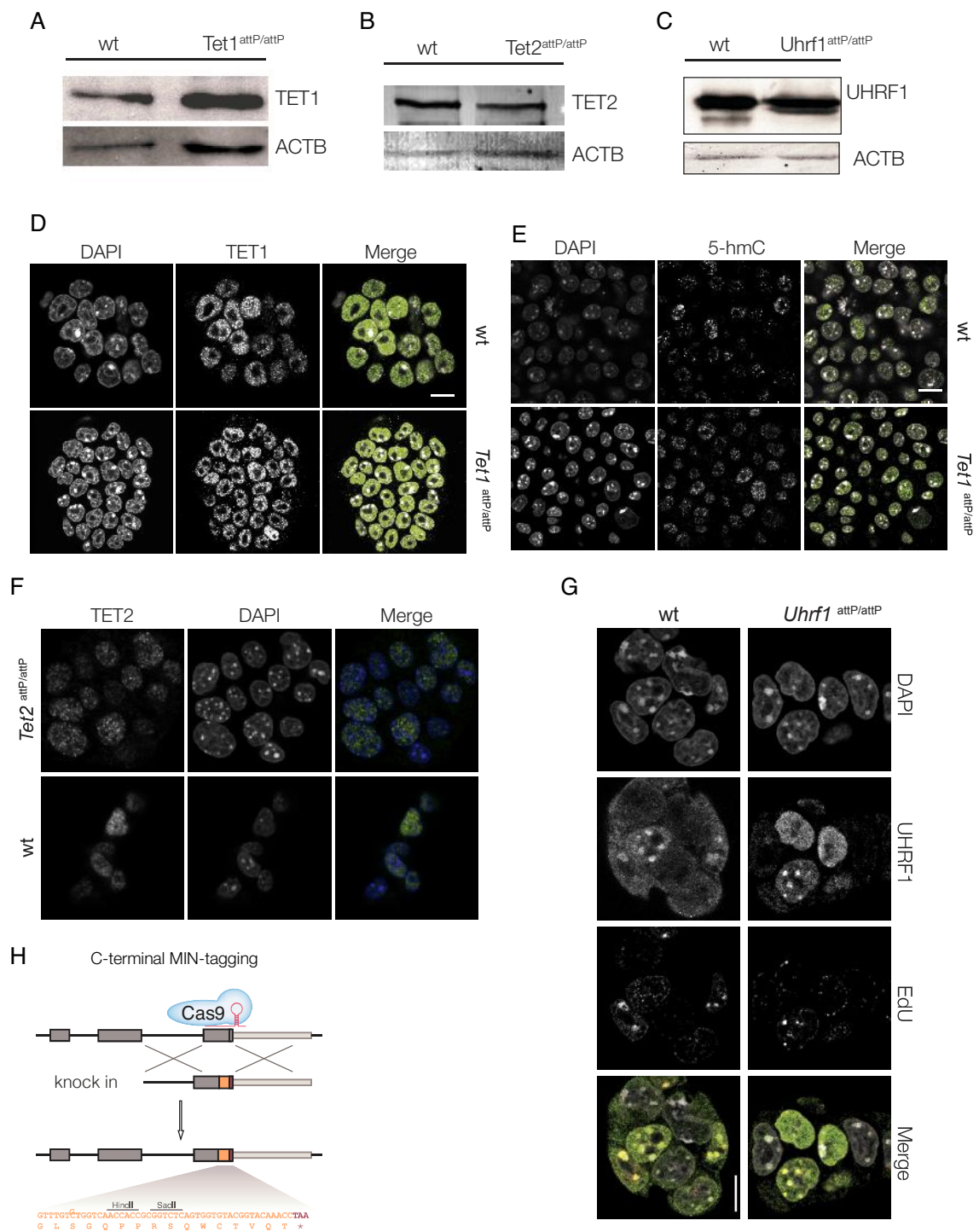
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
- Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. and Doudna, J.A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*, **329**, 1355–1358.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
- Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.-L. and Weiss, D.S. (2013) A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, **497**, 254–257.
- Kuscu, C., Arslan, S., Singh, R., Thorpe, J. and Adli, M. (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.*, **32**, 677–683.
- Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.-J. and Yee, J.-K. (2015) Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.*, **33**, 175–178.
- Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S. *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.*, **32**, 670–676.
- Xu, Z., Thomas, L., Davies, B., Chalmers, R., Smith, M. and Brown, W. (2013) Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol.*, **13**, 87–87.
- Brown, W.R.A., Lee, N.C.O., Xu, Z. and Smith, M.C.M. (2011) Serine recombinases as tools for genome engineering. *Methods*, **53**, 372–379.
- Bonnet, J., Subsoontorn, P. and Endy, D. (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 8884–8889.
- Huang, J., Ghosh, P., Hatfull, G.F. and Hong, Y. (2011) Successive and targeted DNA integrations in the *Drosophila* genome by Bxb1 and phiC31 integrases. *Genetics*, **189**, 391–395.
- Zhu, F., Gamboa, M., Farruggio, A.P., Hippenmeyer, S., Tasic, B., Schule, B., Chen-Tsai, Y. and Calos, M.P. (2014) DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. *Nucleic Acids Res.*, **42**, e34.
- Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P.P. and Bonapace, I.M. (2004) Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol. Cell. Biol.*, **24**, 2526–2535.
- Bauer, C., Gobel, K., Nagaraj, N., Colantuoni, C., Wang, M., Muller, U., Kremmer, E., Rottach, A. and Leonhardt, H. (2015) Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J. Biol. Chem.*, **290**, 4801–4812.
- Schneider, K., Fuchs, C., Dobay, A., Rottach, A., Qin, W., Wolf, P., Alvarez-Castro, J.M., Nalaskowski, M.M., Kremmer, E., Schmid, V. *et al.* (2013) Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. *Nucleic Acids Res.*, **41**, 4860–4876.
- Solovei, I. and Cremer, M. (2010) 3D-FISH on cultured cells combined with immunostaining. *Methods Mol. Biol.*, **659**, 117–126.
- Meilinger, D., Fellinger, K., Bultmann, S., Rothbauer, U., Bonapace, I.M., Klinkert, W.E.F., Spada, F. and Leonhardt, H. (2009) Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep.*, **10**, 1259–1264.
- Niwa, H., Yamamura, K. and Miyazaki, J. (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene*, **108**, 193–199.
- Poser, I., Sarov, M., Hutchins, J.R.A., Heriche, J.-K., Toyoda, Y., Pozniakovskiy, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W. *et al.* (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods*, **5**, 409–415.
- Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Honing, K. and Hornung, V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.*, **31**, 76–81.
- Masui, S., Shimosato, D., Toyooka, Y., Yagi, R., Takahashi, K. and Niwa, H. (2005) An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. *Nucleic Acids Res.*, **33**, e43.
- Easwaran, H.P., Schermelleh, L., Leonhardt, H. and Cardoso, M.C. (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep.*, **5**, 1181–1186.
- Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenoder, S., Wang, M., Qin, W., Soding, J., Spada, F. *et al.* (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS One*, **6**, e16627.
- Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C. and Leonhardt, H. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.*, **35**, 4301–4312.
- Hayashi, K. and Saitou, M. (2013) Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. *Nat. Protoc.*, **8**, 1513–1524.
- Hermann, M., Stillhard, P., Wildner, H., Seruggia, D., Kapp, V., Sanchez-Iranzo, H., Mercader, N., Montoliu, L., Zeilhofer, H.U. and Pelczar, P. (2014) Binary recombinase systems for high-resolution conditional mutagenesis. *Nucleic Acids Res.*, **42**, 3894–3907.

27. Roux,K.J., Kim,D.I., Raida,M. and Burke,B. (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.*, **196**, 801–810.
28. Baymaz,H.I., Spruijt,C.G. and Vermeulen,M. (2014) Identifying nuclear protein-protein interactions using GFP affinity purification and SILAC-based quantitative mass spectrometry. *Methods Mol. Biol.*, **1188**, 207–226.
29. Blum,H., Beier,H. and Gross,H.J. (1987) Improved silver staining of plant-proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis*, **8**, 93–99.
30. Rappsilber,J., Mann,M. and Ishihama,Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*, **2**, 1896–1906.
31. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
32. Cox,J., Hein,M.Y., Luber,C.A., Paron,I., Nagaraj,N. and Mann,M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics*, **13**, 2513–2526.
33. UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
34. Eden,E., Navon,R., Steinfeld,I., Lipson,D. and Yakhini,Z. (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, **10**, 48–48.
35. Kim,H., Um,E., Cho,S.-R., Jung,C., Kim,H. and Kim,J.-S. (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods*, **8**, 941–943.
36. Geiman,T.M., Sankpal,U.T., Robertson,A.K., Chen,Y., Mazumdar,M., Heale,J.T., Schmiesing,J.A., Kim,W., Yokomori,K., Zhao,Y. *et al.* (2004) Isolation and characterization of a novel DNA methyltransferase complex linking DNMT3B with components of the mitotic chromosome condensation machinery. *Nucleic Acids Res.*, **32**, 2716–2729.
37. Lehnertz,B., Ueda,Y., Derijck,A.A.H.A., Braunschweig,U., Perez-Burgos,L., Kubicek,S., Chen,T., Li,E., Jenuwein,T. and Peters,A.H.F.M. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.*, **13**, 1192–1200.
38. Li,E., Bestor,T.H. and Jaenisch,R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
39. Kim,D.I., Birendra,K.C., Zhu,W., Motamedchaboki,K., Doye,V. and Roux,K.J. (2014) Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2453–2461.
40. Tusher,V.G., Tibshirani,R. and Chu,G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 5116–5121.
41. Williams,K., Christensen,J., Pedersen,M.T., Johansen,J.V., Cloos,P.A.C., Rappsilber,J. and Helin,K. (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, **473**, 343–348.
42. Okano,M., Bell,D.W., Haber,D.A. and Li,E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
43. Hata,K., Okano,M., Lei,H. and Li,E. (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*, **129**, 1983–1993.
44. Okano,M., Xie,S. and Li,E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.*, **19**, 219–220.
45. Bernstein,B.E., Kamal,M., Lindblad-Toh,K., Bekiranov,S., Bailey,D.K., Huebert,D.J., McMahon,S., Karlsson,E.K., Kulbokas,E.J., Gingeras,T.R. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**, 169–181.
46. Bachman,K.E., Rountree,M.R. and Baylin,S.B. (2001) Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem.*, **276**, 32282–32287.
47. Xu,G.L., Bestor,T.H., Bourc'his,D., Hsieh,C.L., Tommerup,N., Bugge,M., Hulten,M., Qu,X., Russo,J.J. and Viegas-Pequignot,E. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*, **402**, 187–191.
48. Emperle,M., Rajavelu,A., Reinhardt,R., Jurkowska,R.Z. and Jeltsch,A. (2014) Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. *J. Biol. Chem.*, **289**, 29602–29613.
49. Gowher,H. and Jeltsch,A. (2001) Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *J. Mol. Biol.*, **309**, 1201–1208.
50. Schermelleh,L., Spada,F., Easwaran,H.P., Zolghadr,K., Margot,J.B., Cardoso,M.C. and Leonhardt,H. (2005) Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat. Methods*, **2**, 751–756.
51. Pruitt,K.D., Brown,G.R., Hiatt,S.M., Thibaud-Nissen,F., Astashyn,A., Ermolaeva,O., Farrell,C.M., Hart,J., Landrum,M.J., McGarvey,K.M. *et al.* (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.*, **42**, 756–763.
52. Weisenberger,D.J., Velicescu,M., Cheng,J.C., Gonzales,F.A., Liang,G. and Jones,P.A. (2004) Role of the DNA methyltransferase variant DNMT3b3 in DNA methylation. *Mol. Cancer Res.*, **2**, 62–72.
53. Shalem,O., Sanjana,N.E., Hartenian,E., Shi,X., Scott,D.A., Mikkelsen,T.S., Heckl,D., Ebert,B.L., Root,D.E., Doench,J.G. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, **343**, 84–87.
54. Wang,T., Wei,J.J., Sabatini,D.M. and Lander,E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80–84.
55. Lin,Y., Cradick,T.J., Brown,M.T., Deshmukh,H., Ranjan,P., Sarode,N., Wile,B.M., Vertino,P.M., Stewart,F.J. and Bao,G. (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.*, **42**, 7473–7485.
56. Tan,E.P., Li,Y., Velasco-Herrera Mdel,C., Yusa,K. and Bradley,A. (2015) Off-target assessment of CRISPR-Cas9 guiding RNAs in human iPS and mouse ES cells. *Genesis*, **53**, 225–236.
57. Russell,J.P., Chang,D.W., Tretiakova,A. and Padidam,M. (2006) Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells. *Biotechniques*, **40**, 460–460.
58. Keravala,A., Groth,A.C., Jarrahan,S., Thyagarajan,B., Hoyt,J.J., Kirby,P.J. and Calos,M.P. (2006) A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. *Mol. Genet. Genomics*, **276**, 135–146.
59. Banaszynski,L.A., Chen,L.-C., Maynard-Smith,L.A., Ooi,A.G.L. and Wandless,T.J. (2006) A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell*, **126**, 995–1004.
60. Kennedy,M.J., Hughes,R.M., Peteya,L.A., Schwartz,J.W., Ehlers,M.D. and Tucker,C.L. (2010) Rapid blue-light-mediated induction of protein interactions in living cells. *Nat. Methods*, **7**, 973–975.
61. van Steensel,B., Delrow,J. and Henikoff,S. (2001) Chromatin profiling using targeted DNA adenine methyltransferase. *Nat. Genet.*, **27**, 304–308.
62. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
63. Costa,Y., Ding,J., Theunissen,T.W., Faiola,F., Hore,T.A., Shliha,P.V., Fidalgo,M., Saunders,A., Lawrence,M., Dietmann,S. *et al.* (2013) NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature*, **495**, 370–374.
64. Chen,T., Tsujimoto,N. and Li,E. (2004) The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Mol. Cell Biol.*, **24**, 9048–9058.

Supplementary Figure 1

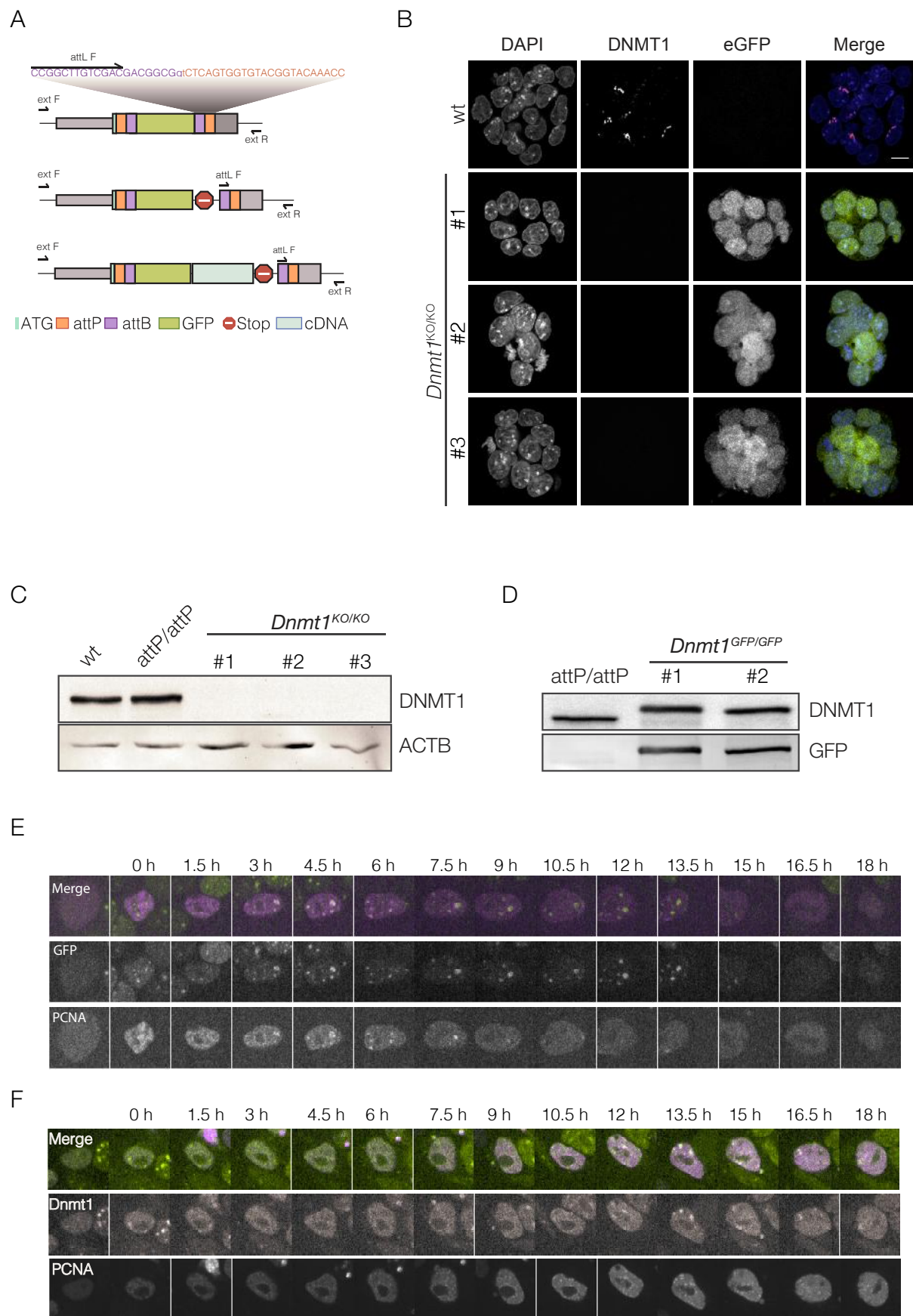


Supplementary Figure 2



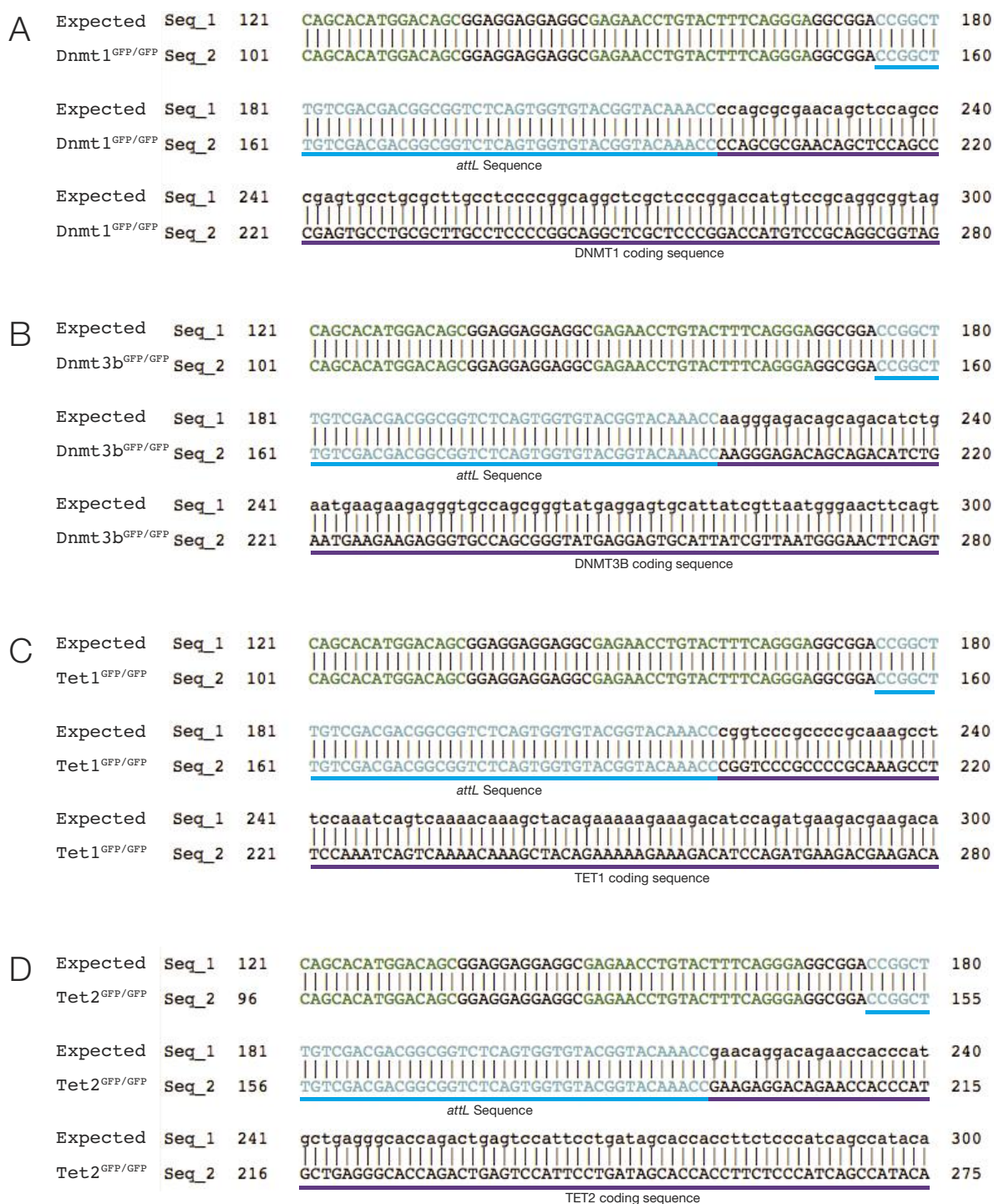


## Supplementary Figure 3

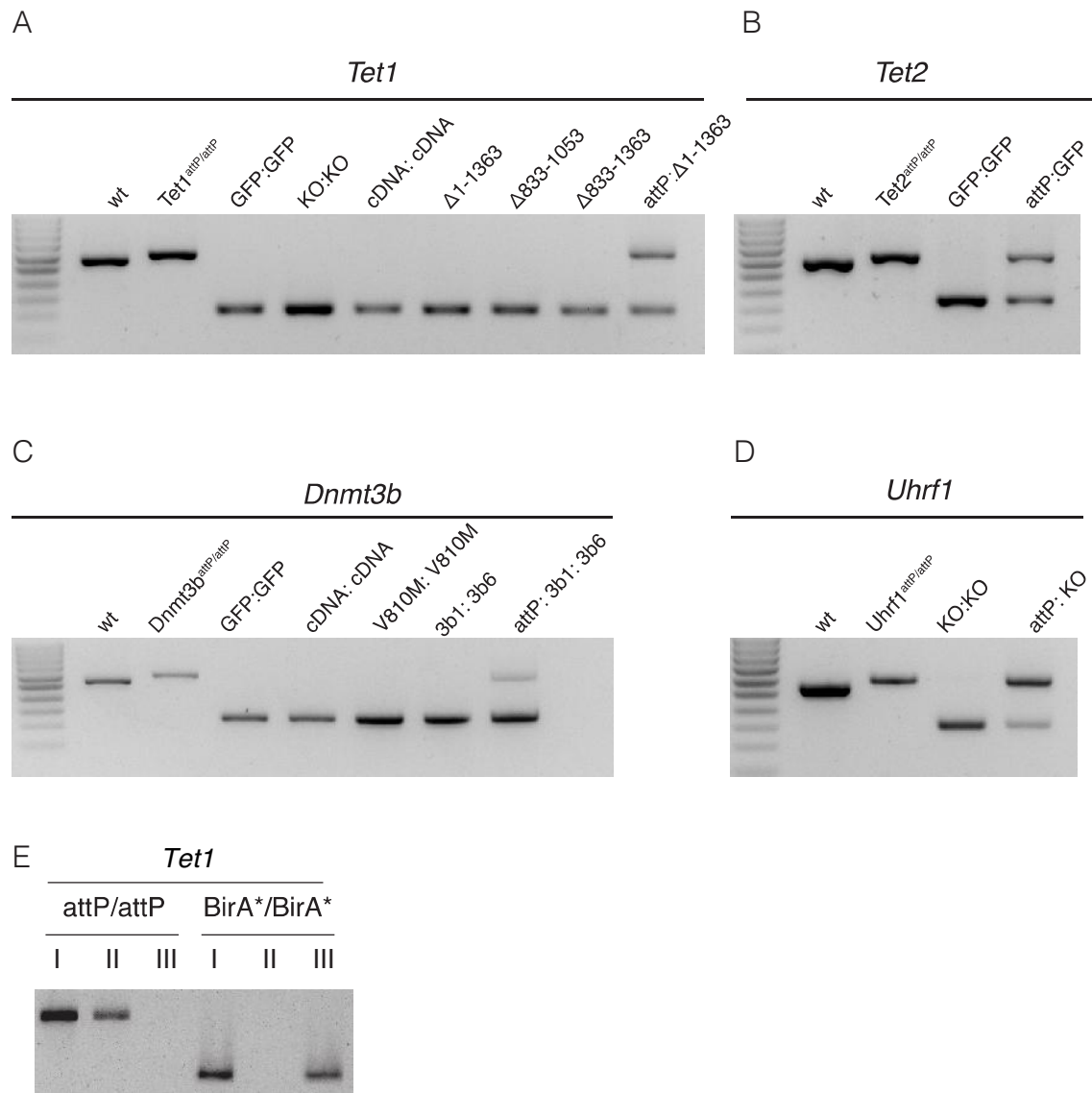




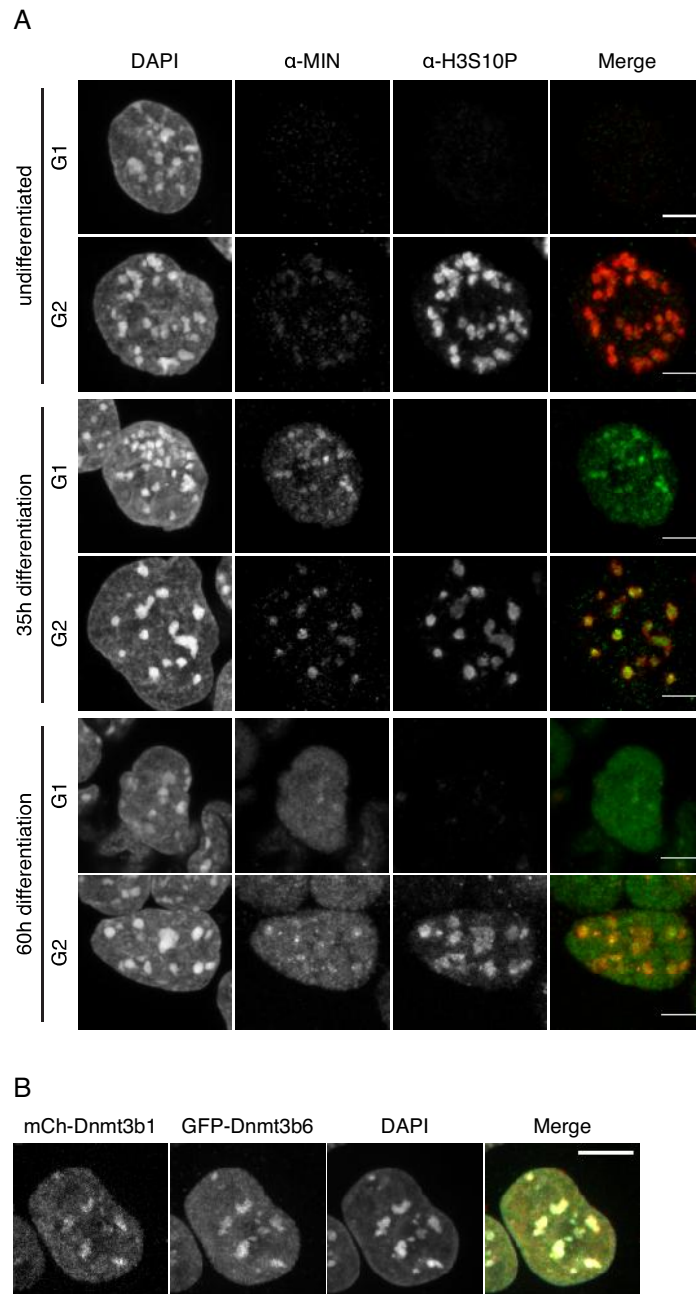
## Supplementary Figure 4



## Supplementary Figure 5



## Supplementary figure 6



## Supplemental Figure and Video Legends

### Supplemental Figure S1. Characterization of MIN-tagged DNA methyltransferase cell lines.

**(A)** Western blot analysis of DNMT1 expression levels in the homozygous *Dnmt1*<sup>attP/attP</sup> and wild type J1 cells. Beta-actin is used as a loading control. **(B)** Immunofluorescence stainings of Dnmt1 in wt and *Dnmt1*<sup>attP/attP</sup> cells. Scale bar represent 5  $\mu$ m. **(C)** DNA methylation analysis of the major satellite repeats in *Dnmt1*<sup>attP/attP</sup> and wild type cells. **(D)** Example of the screening PCRs, with and without HincII treatment, of clones found to be heterozygous and homozygous for the MIN-tag at the *Dnmt1* locus. Monoallelic and biallelic insertions of the MIN-tag can be distinguished by complete and incomplete digests, respectively. **(E)** Western blot analysis of DNMT3A expression levels in a heterozygous (#2) and homozygous (#1, #3-4) *Dnmt3a*<sup>attP/attP</sup> cell lines compared to wild type cells. Beta-actin is used as a loading control. **(F)** DNA methylation analysis of major satellite repeats in *Dnmt3a*<sup>attP/attP</sup> compared to wt cells. **(G)** Immunofluorescence stainings of DNMT3A together with the replication marker EdU in wt cells and the the homozygous *Dnmt3a*<sup>attP/attP</sup> clone #1. Scale bar represents 10  $\mu$ m. **(H)** Immunofluorescence stainings of DNMT3B in *Dnmt3b*<sup>attP/attP</sup> and wt cells after 35 hours of EpiLC differentiation. Scale bar represents 5  $\mu$ m. Error bar represent standard deviation (n=2).

### Supplemental Figure S2. Characterization of MIN-tagged Tet1, Tet2 and Uhrf1 cell lines and C-terminal MIN-tag integration.

**(A-C)** Western blot analysis of TET1, TET2, and UHRF1 expression levels in the homozygous *Tet1*<sup>attP/attP</sup>, *Tet2*<sup>attP/attP</sup>, and N-terminal *Uhrf1*<sup>attP/attP</sup> cell lines, respectively, compared to the wt J1 control.  $\beta$ -Actin (ACTB) was used as loading control. **(D)** Immunofluorescence stainings of TET1 in wt and *Tet1*<sup>attP/attP</sup> cells. **(E)** Immunofluorescence stainings of 5-hydroxymethylcytosine (5-hmC) in wt and *Tet1*<sup>attP/attP</sup> cells. **(F)** Immunofluorescence stainings of TET2 in wt and *Tet2*<sup>attP/attP</sup> cells. **(G)**

Immunofluorescence stainings of UHRF1 in wt and *Uhrf1<sup>attP/attP</sup>* cells. DAPI is used for DNA counterstaining; scale bars represent 15  $\mu$ m. **(H)** Schematic overview of CRISPR/Cas-assisted C-terminal integration of the MIN-tag. MIN-tag donors contain the attP site (depicted in orange) flanked by sequences (200-300 for PCR fragments or 76 for ssDNA oligos) homologous to 5' and 3' of the target gene stop codon (depicted in red). Restriction enzyme sites available for restriction fragment analysis based screening are shown above the attP sequence.

### **Supplemental Figure S3. Evaluating functionality of Bxb1 mediated recombination in *Dnmt1<sup>attP/attP</sup>* cells.**

**(A)** Schematic outline of the multiplex PCR strategy to identify positive recombination events and their zygosity. **(B)** Immunofluorescence stainings of DNMT1 and GFP in wt cells and three *Dnmt1<sup>KO/KO</sup>* clones. Diffuse GFP indicates a successful integration of the KO cassette into the locus. **(C)** Western blot analysis of DNMT1 expression levels in three *Dnmt1<sup>KO/KO</sup>* clonal cell lines generated by Bxb1-mediated insertion of a knock-out cassette, compared to wt and *Dnmt1<sup>attP/attP</sup>* cells. **(D)** Western blot analysis of DNMT1 and GFP expression in *Dnmt1<sup>attP/attP</sup>* cells and two homozygous GFP-knock in cell lines (*Dnmt1<sup>GFP/GFP</sup>* #1-2) generated by Bxb1 mediated insertion. **(E-F)** Live cell imaging of *Dnmt1<sup>GFP/GFP</sup>* and *Dnmt1<sup>cDNA/cDNA</sup>* cells transiently expressing RFP-labeled PCNA, a DNA replication marker, during cell-cycle progression. Scale bars represent 5  $\mu$ m

### **Supplemental Figure S4. Alignments of the expected sequence flanking the attL site after recombination**

Alignments of the expected sequence flanking the attL site after recombination of the attB-GFP KI at the *Dnmt1*, *Dnmt3b*, *Tet1*, and *Tet2* locus (A-D) with the sequencing results from the *Dnmt1<sup>GFP/GFP</sup>*, *Dnmt3b<sup>GFP/GFP</sup>*, *Tet1<sup>GFP/GFP</sup>*, and *Tet2<sup>GFP/GFP</sup>* cell lines.

### **Supplemental Figure S5. Demonstration of Bxb1 mediated recombination in multiple MIN-tagged genes.**

**(A-D)** Gel electrophoresis of the multiplex PCR (using the attL primer and locus specific

external primers, see also Table S1) performed on cell lines generated by Bxb1-mediated integration of various MIN-tag toolbox components (Table S5) into the loci of: (A) Tet1, (B) Tet2, (C) Dnmt3b, and (D) Uhrf1. Equal mixtures of genomic DNA from non-recombined cell lines and recombined cell lines are used to control for possible amplification biases arising from the use of different locus specific external primers. **(E)** PCR to confirm insertion of the BirA\* cassette into the Tet1 genomic locus. I: multiplex PCR, II: wt specific PCR, III: attL (recombination) specific PCR

**Supplemental Figure S6. Cell cycle analysis of DNMT3b localization during differentiation.**

**(A)** Immunofluorescence stainings of MIN-tagged DNMT3B and Histone 3 Serine 10 phosphorylation (H3S10P), a marker of G2/M phase (Ref Hendzel:1997wo) during differentiation of naive pluripotent *Dnmt3b*<sup>attP/attP</sup> stem cells into epiblast-like cells. Cells were fixed directly after (0 h) 35 h, or 60 h after induction of differentiation. The H3S10P mark was used to determine if cells were in G2 or G1 phase in order to assess whether changes in DNMT3B localization during differentiation are cell-cycle dependent. Scale bar represents 5  $\mu$ m. **(B)** Fluorescence microscopy images of *Dnmt3b*<sup>mCh-3b1/GFP-3b6</sup> cells fixed after 35 h of differentiation. Both DNMT3B isoforms (GFP-DNMT3B1 in green and mCh-DNMT3B6 in red) localize at chromocenters (visible as bright DAPI spots). Scale bar represents 5  $\mu$ m

**Supplemental Video 1. Live cell imaging of *Dnmt3b*<sup>GFP/GFP</sup> cells during differentiation.**

Long-term (60 h), live cell imaging tracking the transition of *Dnmt3b*<sup>GFP/GFP</sup> cells from the naive pluripotency ground state into the primed, epiblast-like state. Images were acquired once per hour and entailing at least 10  $\mu$ m z-stacks. The left panel depicts the projection of GFP signal, while the right panel shows that projection superimposed onto the acquired brightfield images.



## Supplemental Tables (S1-S5)

**Table S1: CRISPR/Cas9-mediated MIN-tag insertion efficiencies**

Gene	Position	MIN-tag Donor	Heterozygotes	Homozygotes	TOTAL
<b>Dnmt1</b>	N-terminal	PCR Product	1/67 (1.5%)	1/67 (1.5%)	2/67 (2.9%)
<b>Dnmt3a</b>	N-terminal	PCR Product	0/86 (0%)	3/86 (3.5%)	3/86 (3.5%)
<b>Dnmt3b</b>	N-terminal	ssDNA oligo	0/65(0%)	1/65(1.5%)	1/65 (1.5%)
<b>Uhrf1</b>	N-terminal	PCR Product	0/6 (0%)	1/6(16.7%)	1/6 (16.7%)
<b>Uhrf1</b>	C-terminal	ssDNA oligo	2/36 (5.5%)	2/36 (5.6 %)	4/36 (11.1%)
<b>Tet1</b>	N-terminal	PCR Product	0/70(0%)	1/70 (1.4%)	1/70 (1.4%)
<b>Tet2</b>	N-terminal	PCR Product	1/24 (4.2%)	2/24 (8.3%)	3/24 (12.5%)
<b>Tet3</b>	N-terminal	PCR Product	0/38 (0%)	2/38 (5.3%)	2/38 (5.3%)

**Table S2: Oligonucleotide sequences used for CRISPR/Cas assisted targeting and screening**

Name	Sequence
Dnmt1	
gRNA_F	TGTTTCGCGCTGGCATCTTGCCTTTTAGAGCTAGAAATAGCAAG
gRNA_R	GCAAGATGCCAGCGCGAACACGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCTGTTTCGCGCTGGCATCTTGCAGGGGATTCC
surrogate_R	CCGGAGGAATCCCCCTGCAAGATGCCAGCGCGAACAG
internal_R	CACTATAGCCAGGAGGTGTGGG
internal_F	TGTACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCCATCTTGCAGGTTGCA GACGACAG
external_R	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCAGCGCGAACAGCTCC AGC
external_F	GCGCGACAGGAAGCACAGCC
screening_F	GTTCGACGACGGACGAG
Uhrf1 (N)	
gRNA_F	CATCGGCATCATGTGGATCCGTTTATAGAGCTAGAAATAGCAAG
gRNA_R	GGATCCACATGATGCCGATGCGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCCATCGGCATCATGTGGATCCAGGGGATTCTT
surrogate_R	GGCCAGGAATCCCCCTGGATCCACATGATGCCGATGG
internal_R	CATCGGCATCATGTGGATCCGTTTATAGAGCTAGAAATAGCAAG
internal_F	GGATCCACATGATGCCGATGCGGTGTTTCGTCCTTTCCAC
external_R	ACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGGATCCAGGTTCTGAACCTATG
external_F	CTATTGCTTGGTGGCTTTGAG
screening_F	GGCAATTCACATTCAAGTGTCCC
Uhrf1 (C)	
gRNA_F	TGCCTGGGTCTCAGCATCACGTTTATAGAGCTAGAAATAGCAAG
gRNA_R	GTGATGCTGAGACCCAGGCACGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCTGCCTGGGTCTCAGCATCACCGGGGATTCTT
surrogate_R	CCGGAGGAATCCCCGGTGATGCTGAGACCCAGGCAG
ssDNA oligo	CAGCTCCCCAACCCGGGTGAACCAGCCCTTGCAGACCATTCTCAACCAGCTCTTCCCTGG CTATGGCAGCGGCCGGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACA AACCTGATGCTGAGACCCAGGCAGAGGGCTCATGGTTCCAACCTTCATAGTGTGTTTAGCT TGAAGGTGTTGTCTTCACG
external_R	TTTCTAGGCAGCTGGTGTGG
external_F	TGTACGTGAGAGGACGGAGT
screening_F	TGTTGCCAGGAGCTACCAAG
Dnmt3a	
gRNA_F	GGGCCGCTGGAGGGCATTGCGTTTATAGAGCTAGAAATAGCAAG
gRNA_R	GCAATGCCCTCCAGCGGCCCGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCGGGCCGCTGGAGGGCATTGCTGGGGATTCTT
surrogate_R	CCGGAGGAATCCCCAGCAATGCCCTCCAGCGGCCCG
internal_R	CTTCTCTTCCCCACAGGCAG
internal_F	ACCACTGAGACCGCGGTGGTTGACCAGACAAACCCATTGCTGGGCAGTAGGCG
external_R	ACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCCTCCAGCGGCCCG
external_F	GTTCCCAGCCAAGCACCTAT
screening_F	ATGGTCCTGCAACCAGAGTG
Dnmt3b	
gRNA_F	TTCCCCACAGGAAACAATGAGTTTATAGAGCTAGAAATAGCAAG
gRNA_R	TCATTGTTTCCTGTGGGGAACGGTGTTTCGTCCTTTCCAC

## 2. Results

surrogate_F	CTAGCTTCCCCACAGGAAACAATGAAGGGGATTCCCT
surrogate_R	CCGGAGGAATCCCCTTCATTGTTTCCTGTGGGGAAG
ssDNA oligo	GAAGTGGTGGTGTAAACCTTGCAGTGTGCCCTGTCTGCCTCTTACATATCCTGATCTTTC
	CCCACAGGAAACAATGGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACA
external_R	AACCAAGGGAGACAGCAGACATCTGAATGAAGAAGAGGGTGCCAGCGGGTATGAGGAGTG
	CATTATCGTTAATGGGAACT
external_F	ACCACCGCGTCTCAGTGGTGTACGGTACAAACCGGAGACAGCAGACATCTGAATG
screening_F	ATCTGTCATGGAACCTGCCG
	GAGCTGGCCAATTGCAGAAC
Tet1	
gRNA_F	AGACATGGCTGCAGAGTAAGCGGTGTTTCGTCCTTTCCAC
gRNA_R	CTTACTCTGCAGCCATGTCTAGCTTTCTTGTACAAAGTTGGCAT
surrogate_F	CTAGCCTTACTCTGCAGCCATGTCTCGGGGATCCCT
surrogate_R	CCGGAGGGATCCCCGAGACATGGCTGCAGAGTAAGG
internal_R	ACTCAGTCTCCCAAATGCTGG
internal_F	ACCACTGAGACCGCGGTGGTTGACCAGACAAACCAGACATGGCTGCAGAGTAAGTAAAG
external_R	ACCACCGCGTCTCAGTGGTGTACGGTACAAACCGGTCCCGCCCCGCAAAG
external_F	TCGGGGTTTTGTCTTCCGTT
screening_F	GGGCAATGTTGTGACTCATGC
Tet2	
gRNA_F	CGAAGCAAGCCTGATGGAACGTTTTAGAGCTAGAAATAGCAAG
gRNA_R	GTTCCATCAGGCTTGCTTCGCGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCCGAAGCAAGCCTGATGGAACAGGGGATTCCCT
surrogate_R	CCGGAGGAATCCCCTGTTCCATCAGGCTTGCTTCGG
internal_R	ACCACTGAGACCGCGGTGGTTGACCAGACAAACCCATCAGGCTTGCTTCGGGG
internal_F	ACCACCGCGTCTCAGTGGTGTACGGTACAAACCGAACAGGACAGAACCACCCAT
external_R	TGGTTCACCTGACTGTGCGTT
external_F	CCAGGATCACACAGGAAGCA
screening_F	GGATGGAGCCCAGAGAGAGA
Tet3	
gRNA_F	GTTCCAGGTCAGATGGACTCGTTTTAGAGCTAGAAATAGCAAG
gRNA_R	GAGTCCATCTGACCTGGAACCGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCGTTCCAGGTCAGATGGACTCAGGGGATTCCCT
surrogate_R	CCGGAGGAATCCCCTGAGTCCATCTGACCTGGAACG
internal_R	ACCACTGAGACCGCGGTGGTTGACCAGACAAACCCATCTGACCTGGAACAGGTC
internal_F	ACCACCGCGTCTCAGTGGTGTACGGTACAAACCGACTCAGGGCCAGTGTACC
external_R	CAGTCGGGCTTCTGGTCTAC
external_F	GATCTGAGCTCTCACAGGGC
screening_F	AGTAGACAGGGCCTTGGGAT
attL_F	CCGGCTTGTCGACGACG

**Table S3: Bxb1-mediated recombination efficiencies**

Gene	Integration Construct	Heterozygotes	Homozygotes	TOTAL
<b>Dnmt1</b>	attB-GFP	N/A	13/31 (41.9%)	13/31 (41.9%)
<b>Dnmt3b</b>	attB-GFP	0/3 (0%)	1/3 (33.3%)	1/3 (33.3%)
<b>Tet1</b>	attB-GFP	14/45 (31.1%)	13/45 (28.9%)	27/45 (60%)
<b>Tet2</b>	attB-GFP	28/81 (34.6%)	15/81 (18.5%)	43/81 (53%)
<b>Dnmt1</b>	attB-GFP-STOP-Poly(A)	2/23 (8.7%)	13/23 (56.5%)	15/23 (65.2%)
<b>Uhrf1</b>	attB-GFP-STOP-Poly(A)	5/32 (15.6%)	14/32 (43.8%)	19/32 (59.4%)
<b>Dnmt1</b>	attB-GFP-cDNA-STOP-Poly(A)	1/15 (6.6%)	9/15 (60%)	10/15 (66.6%)
<b>Dnmt3b</b>	attB-GFP-cDNA-STOP-Poly(A)	28/84 (33.3%)	26/84 (31%)	54/84 (64.3%)
<b>Tet1</b>	attB-GFP-cDNA-STOP-Poly(A)	12/58 (20.7%)	7/58 (12.1%)	19/58 (32.8%)
<b>Dnmt3b</b>	attB-GFP/mCh-cDNA-STOP-Poly(A) PuroR/neoR	29/102 (28.4%)	64/102 (62.7%)	93/102 (91.2%)

**Table S4: Evaluation of FRAP protein kinetics**

	GFP-DNMT3B	mCh-DNMT3B1	GFP-DNMT3B6
<b>Mobile fraction [A]</b>	87	81	100
<b>Diffusion coef. [<math>\mu\text{m}^2/\text{s}</math>]</b>	4.2E-03	1.2E-03	4.1E-02
<b>Half-time recovery [s]</b>	42.2	94.8	5.1

**Table S5: The MIN-tag toolbox**

Name	Fluorescent protein	Application
<b>Universal constructs</b>		
attB-GFP	GFP	GFP KI
attB-mCh	mCherry	mCherry KI
attB-GFP-T2A-BirA*	GFP	Protein interaction
attB-GFP-Poly(A)	GFP	KO
attB-mCh-Poly(A)	mCherry	KO
attB-GFP-Poly(A)-NeoR	GFP	KO /w selection
attB-GFP-Poly(A)-PuroR	GFP	KO /w selection
attB-mCh-Poly(A)-NeoR	mCherry	KO /w selection
attB-mCh-Poly(A)-PuroR	mCherry	KO /w selection
<b>Gene specific cDNA KI constructs</b>		
attB-GFP-Dnmt1-Poly(A)	GFP	cDNA KI
attB-GFP-Dnmt3b1-Poly(A)	GFP	cDNA KI
attB-GFP-Dnmt3b6-Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_C656A_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_D809G_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_dX_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_G655S_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_L656T_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_V718G_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_V810M_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b6_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b1_dPWWP_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b1_dPHD_Poly(A)	GFP	cDNA KI
attB_mCh_Dnmt3b_C656A_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_D809G_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_dX_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_G655S_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_L656T_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_V718G_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_V810M_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b6_Poly(A)	mCherry	cDNA KI
attB-GFP-Dnmt3b1-Poly(A) -NeoR	GFP	cDNA KI /w selection
attB-GFP-Dnmt3b6-Poly(A) -NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_C656A_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_D809G_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_dX_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_G655S_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_L656T_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_V718G_Poly(A)-NeoR	GFP	cDNA KI /w selection

attB_eGFP_Dnmt3b_V810M_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b6_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB- mCh -Dnmt3b1-Poly(A) -NeoR	mCherry	cDNA KI /w selection
attB- mCh -Dnmt3b6-Poly(A) -NeoR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_C656A_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_D809G_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_dX_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_G655S_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_L656T_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_V718G_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_V810M_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b6_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB- mCh -Dnmt3b1-Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB- mCh -Dnmt3b6-Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_C656A_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_D809G_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_dX_Poly(A)- PuroR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_G655S_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_L656T_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_V718G_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b_V810M_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_ mCh _Dnmt3b6_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB-GFP-Tet1-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d1-389-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d390-565-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d566-833-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d834-1053-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d1054-1363-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d1-833-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d834-1363-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d1-225-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d226-398-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d399-650-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d651-848-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d849-1038-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d1-650-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d651-1038-Poly(A)	GFP	cDNA KI
attB-GFP-Uhrf1-Poly(A)	GFP	cDNA KI
attB-GFP-Uhrf1dSRA-Poly(A)	GFP	cDNA KI



## Supplemental Table Legends

### **Table S1: CRISPR/Cas9-mediated MIN-tag insertion efficiencies**

For MIN-tag Insertion, J1 mESCs transfected with the appropriate MIN-tag donor oligonucleotides or PCR products along with the Cas9, gRNA, and CRISPR surrogate reporter vector were single cell sorted after enriching for cells with CRISPR/Cas activity. The number of clones with either a monoallelic or biallelic insertion of the MIN-Tag is shown in relation to the number of clones screened.

### **Table S2: Oligonucleotide sequences used for CRISPR/Cas assisted targeting and screening**

DNA oligonucleotides used for the generation of target specific gRNA expression vectors, surrogate reporters, and homology donors for MIN-tag integration.

### **Table S3: Bxb1-mediated recombination efficiencies**

For Bxb1-mediated recombination, J1 mESCs transfected with NLS-Bxb1, the Bxb1 surrogate reporter, and the respective attB-site containing integration construct were single-cell sorted after enrichment for cells with Bxb1 activity. The number of clones with either a monoallelic or biallelic integration of the listed construct is shown in relation to the total number of clones screened.

### **Table S4: Evaluation of FRAP protein kinetics**

Evaluation of FRAP kinetics (w/o 5-azadC treatment) performed in Dnmt3bGFP/GFP and Dnmt3bmCh-3b1/GFP-3b6 cells

### **Table S5: The MIN-tag toolbox**

Vectors generated for Bxb1 mediated recombination into MIN-tagged cell lines. KO: knockout, KI: knockin

## **2.5 The impact of DNA modifications on chromatin composition in embryonic stem cells and epiblast-like cells**

unpublished manuscript

## The impact of DNA modifications on chromatin composition in embryonic stem cells and epiblast-like cells

Christina Bauer <sup>a</sup>, Carina Trummer <sup>a</sup>, Udo Müller <sup>a</sup>, Christopher B. Mulholland <sup>a</sup>, Nagarjuna Nagaraj <sup>b</sup>, Heinrich Leonhardt <sup>a,c\*</sup>

<sup>a</sup> LMU Munich, Biology II, Planegg-Martinsried, Germany

<sup>b</sup> Max Planck Institute for Biochemistry, Martinsried, Germany

<sup>c</sup> Center for Integrated Protein Science Munich (CIPSM), Germany

\* to whom correspondence should be addressed: Prof. Dr. Heinrich Leonhardt, Ludwig-Maximilians University Munich, Department Biologie II, Großhadernerstr. 2, 81925 Planegg-Martinsried, Germany Tel: +49 (0) 89 / 2180 - 74229, Fax: +49 (0) 89 / 2180 - 74236, h.leonhardt@lmu.de

### Abstract

DNA modifications play a crucial role in epigenetic regulation in mammals und dynamically change during early embryonic development. However, the interplay between DNA modifications and chromatin proteins is complex and still barely understood. Here, we differentiate embryonic stem cells (ESCs) to epiblast-like cells (EpiLCs), a cell culture model that recapitulates peri-implantation development, and investigate the total chromatin composition by proteomics. We furthermore analyse knockout (KO) cell lines that lack the major DNA modifying enzymes, DNMTs, TETs, and TDG, and describe how loss of these enzymes influences the chromatin proteome. We identify several proteins that are characteristic for chromatin of ESCs, e.g. KLF4, TET2, and TFCP2L1, and for EpiLC chromatin, e.g. DNMT3A, DNMT3B, and SALL2. Finally, we show that transcription factors like KLF4, OCT4, or NonO are depleted from chromatin in ESCs deficient for DNMTs, and that proteins of various DNA repair pathways accumulate at the chromatin of *Tdg*-KO cells.

## Introduction

Epigenetic mechanisms constitute an additional layer of information beyond the DNA sequence, regulating genome architecture and gene expression. Besides modifications of histone tails, one of the best studied epigenetic marks in mammals is cytosine methylation, a process catalysed by DNA methyltransferases (DNMTs)<sup>1</sup>. Oxidative derivatives of methylcytosine (mC) — hydroxymethylcytosine (hmC), formylcytosine (fC) and carboxylcytosine (caC) — have been described for decades<sup>2</sup>, but only recently were discovered to be generated through a catalytic step by TET dioxygenases<sup>3-5</sup>. During early embryonic development, levels of these DNA modifications undergo dramatic changes, from an initial rapid loss of mC in the paternal pronucleus directly after fertilization to a stark global increase of genomic methylation during implantation<sup>6</sup>. The demethylation of the paternal pronucleus is accompanied by TET3-mediated formation of hmC<sup>7</sup>, however, the importance of this step is still debated<sup>8</sup>. The gain of cytosine methylation during implantation, when cells of the inner cell mass (ICM) develop into the epiblast, depends on the activity of the *de novo* methyltransferases DNMT3A and DNMT3B<sup>9</sup>.

In addition to DNMT3A and DNMT3B, major enzymes involved in DNA methylation and demethylation include the maintenance DNA methyltransferase DNMT1, the three TET dioxygenases TET1, TET2, and TET3, and the G-T-mismatch glycosylase TDG, which excises fC and caC with high affinity<sup>3,4</sup>. In general, DNA methylation is associated with transcriptional repression, and affects repetitive elements, imprinting control regions, and the inactive X chromosome<sup>1</sup>. Oxidation of mC by TET proteins and subsequent excision of fC and caC by TDG has been proposed as a mechanism for active DNA demethylation. In line with this hypothesis, fC and caC accumulate at gene regulatory elements in embryonic stem cells (ESCs) deficient for *Tdg*, suggesting a dynamic turnover of cytosine modifications at these sites<sup>10</sup>. However, the oxidized cytosine derivatives might also function as independent epigenetic marks. E.g., hmC has been mapped to gene bodies of transcriptionally active genes and promoter elements of silenced genes<sup>11</sup> and seems to be a stable DNA base generated in a cell cycle-dependent manner<sup>12, 13</sup>. Similarly, a fraction of genomic fC has a low turnover rate, indicating an epigenetic role beyond demethylation<sup>14</sup>.

Cellular interpretation of epigenetic marks is widely thought to be achieved through so-called reader proteins that bind to the mark and ultimately lead to changes in the transcriptional state. For mC, a well-described family of reader proteins has been known since the 1990s, the MBD (methyl CpG binding domain) proteins<sup>15-17</sup>. More recently, proteins that specifically bind to C, mC, hmC, fC, or caC have been identified by a pull-down approach in ESCs, neuronal progenitor cells, and brain tissue<sup>18</sup>. However, chromatin proteins often bind to several epigenetic marks such as histone and DNA modifications in a cooperative manner, and the complex interplay between DNA modifications and proteinaceous chromatin components remains barely understood.

## 2. Results

Here, we aimed to estimate the impact of genomic cytosine modifications on chromatin composition by analysing the chromatin-associated proteins in KO cell lines lacking the major enzymes involved in the DNA methylation and demethylation pathways. To identify and quantify the proteinaceous chromatin components dependent on the presence of DNA modifications, we performed chromatin enrichment and subsequent LC-MS/MS analysis on wild type (wt) ESCs, *Dnmt*-triple-KO (*Dnmt*-TKO), *Tet*-triple-KO (*Tet*-TKO), and *Tdg*-KO cell lines. Furthermore, we employed a cell culture model that recapitulates development from the ICM to the epiblast by differentiating naïve pluripotent ESCs to primed epiblast-like cells (EpiLCs)<sup>19</sup> to investigate changes in chromatin composition at this crucial developmental step. Our dataset serves as a valuable resource to understand the interplay between DNA modifications and chromatin proteins both in ESCs and EpiLCs.

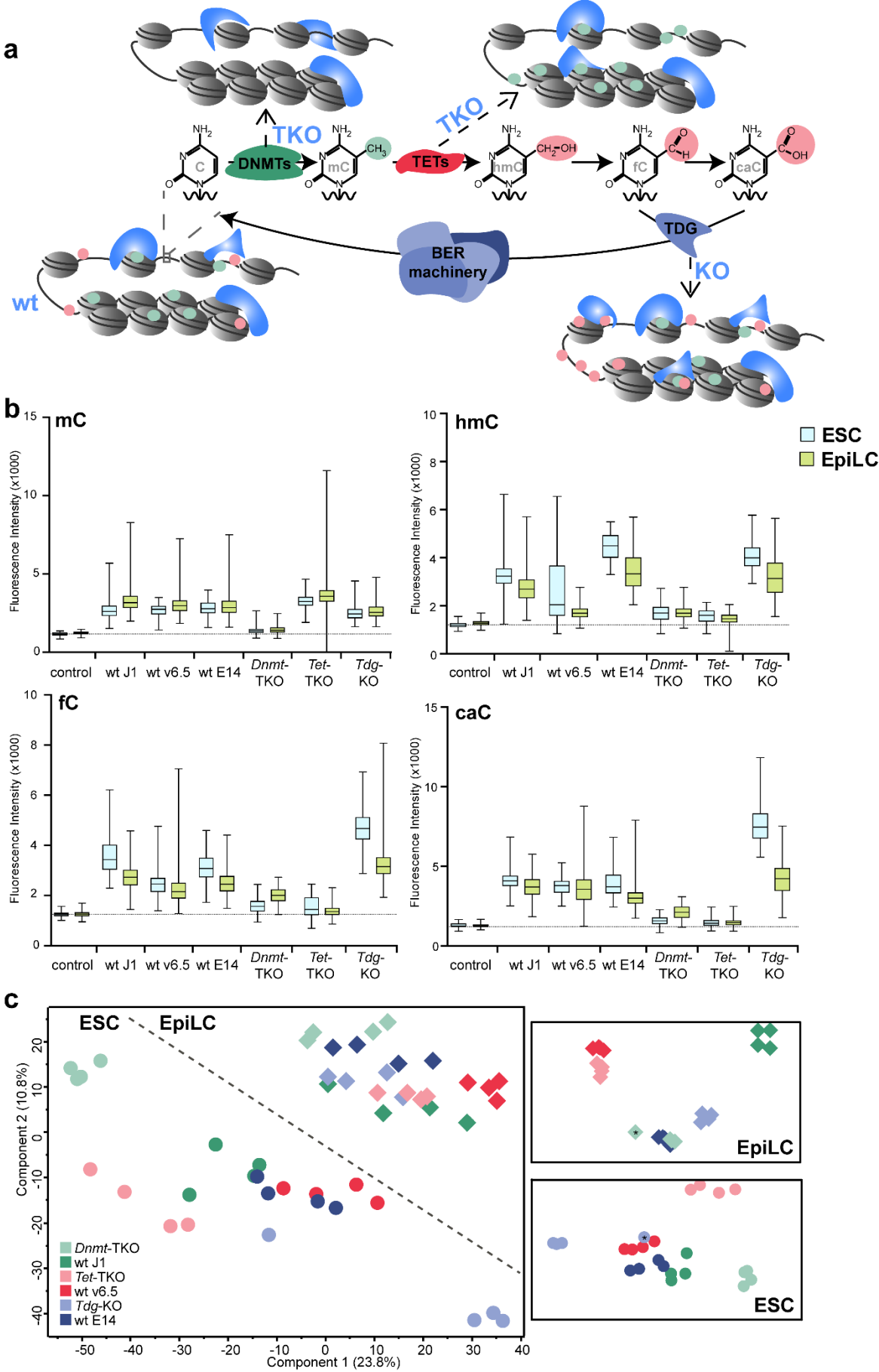
## Results

### Chromatin composition in ESCs and EpiLCs

Genomic DNA is marked by a variety of DNA modifications, mostly targeting cytosine. Loss of the enzymes of the DNA modification pathways has strong effects on the DNA modification landscape. First, KO of all catalytically active DNMTs leads to DNA completely devoid of all cytosine modifications<sup>20</sup>. Second, loss of the three TET enzymes results in accumulation of mC and elimination of the oxidized mC-derivatives hmC, fC, and caC<sup>21</sup>. Finally, KO of the DNA glycosylase TDG<sup>22</sup> causes an increase of genomic fC and caC levels<sup>10, 23</sup> (Fig. 1a,b). To investigate the influence of DNA modifications on chromatin composition, we performed chromatin enrichment for proteomics (ChEP)<sup>24</sup> in *Dnmt*-TKO, *Tet*-TKO, and *Tdg*-KO ESCs and in their corresponding wt cell lines J1, v6.5, and E14. Additionally, we analysed the chromatin of these cell lines after 64 h of EpiLC differentiation.

Morphologically, all six ESC lines can undergo EpiLC differentiation for 64 h (Supplementary Fig. 1). Per cell line and differentiation state, ChEP experiments were performed in quadruplicates, resulting in a total of 48 analysed samples. Principle component analysis (PCA) of the samples reveals a clear separation of the ESC and EpiLC state. In ESCs, the three wt cell lines cluster closely together, indicating a rather similar chromatin composition. The three different KO cell lines form distinct groups separate from the wt and from each other. Chromatin composition in the EpiLC state is more similar among cell lines. Performing a separate PCA for ESCs and EpiLCs shows clustering of the experimental replicates, but not of the three wt samples in EpiLCs, suggesting a larger heterogeneity between cell lines than in ESCs (Fig. 1c).

Figure 1





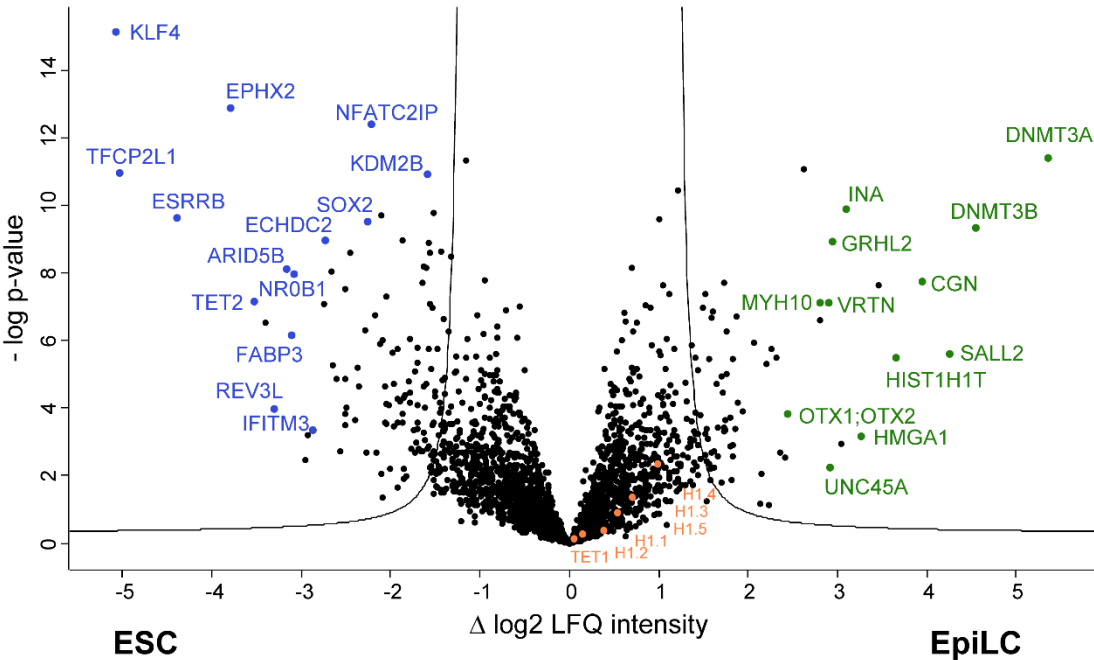
2. Results

**Figure 1: Distinct chromatin composition of ESCs, EpiLCs, and KO cell lines for DNA-modifying enzymes**

- a) Schematic representation of the DNA methylation and demethylation pathway. KO of the participating enzymes results in altered DNA modification patterns and different chromatin composition as illustrated.
- b) Quantification of mC, hmC, fC, and caC by automated high-throughput imaging. Fluorescence intensity reflects the nuclear signal of the respective antibody staining. Note that accumulation of fC and caC in *Tdg*-KO cells is much more pronounced in the ESC state compared to the EpiLC state. Examples of images used for quantification are provided in Supplementary Figure 2. Control: secondary antibody only
- c) PCA of all 48 ChEP samples based on log<sub>2</sub> LFQ (label-free quantification) intensities of 3159 proteins total. circles: ESCs, diamonds: EpiLCs, colours indicate different cell lines, individual data points represent experimental replicates. Boxes: separate PCA of ESC and EpiLC datasets. 2 outliers can be identified among the replicates: *Tdg*-KO ESC replicate 4, and *Dnmt*-TKO EpiLC replicate 2 (marked by asterisks).

After stringent data filtering (for details, see Methods section), 2279 proteins were identified and quantified by ChEP and LC-MS/MS in wt cell lines. Comparison of the chromatin-associated proteins in the EpiLC and the ESC state leads to the identification of 188 significantly different proteins, with 121 specific for ESCs and 67 specific for EpiLCs (Fig. 2). Providing a proof-of-concept for the applied approach, known pluripotency factors like ESRRB or KLF4 are highly abundant in the ESC ChEP, whereas DNMT3A and DNMT3B, the two *de novo* methyltransferases which are upregulated during EpiLC development<sup>19</sup>, are prominently present in the EpiLC samples. Interestingly, TET2 is specific for ESCs ( $\Delta \log_2 \text{LFQ} = -3.52$ , p-value =  $7 \times 10^{-8}$ ), whereas TET1 levels do not change during EpiLC differentiation ( $\Delta \log_2 \text{LFQ} = 0.08$ , p-value = 0.68). A full list of all significantly different proteins is provided as Supplementary Table 1.

**Figure 2**



**Figure 2: Differences in chromatin proteins between wt ESCs and wt EpiLCs**

Volcano plot of wt ESC chromatin versus wt EpiLC chromatin. The three different wt strains in quadruplicates were treated as replicates, resulting in  $n = 12$  for each sample set. Each dot represents one identified protein. Lines indicate the FDR threshold with randomization-based FDR  $< 0.01$  and  $S0 = 2$ ; p-value on the y-axis is based on a student's t-test. A positive  $\Delta \log_2$  LFQ intensity signifies higher abundance of the protein in the EpiLC dataset whereas a negative value means higher abundance in the ESC samples. For reasons of clarity, not all data points are specifically labelled; for a full list see Supplementary Table 1.

**Chromatin proteins of ESCs correlating with presence of DNA modifications**

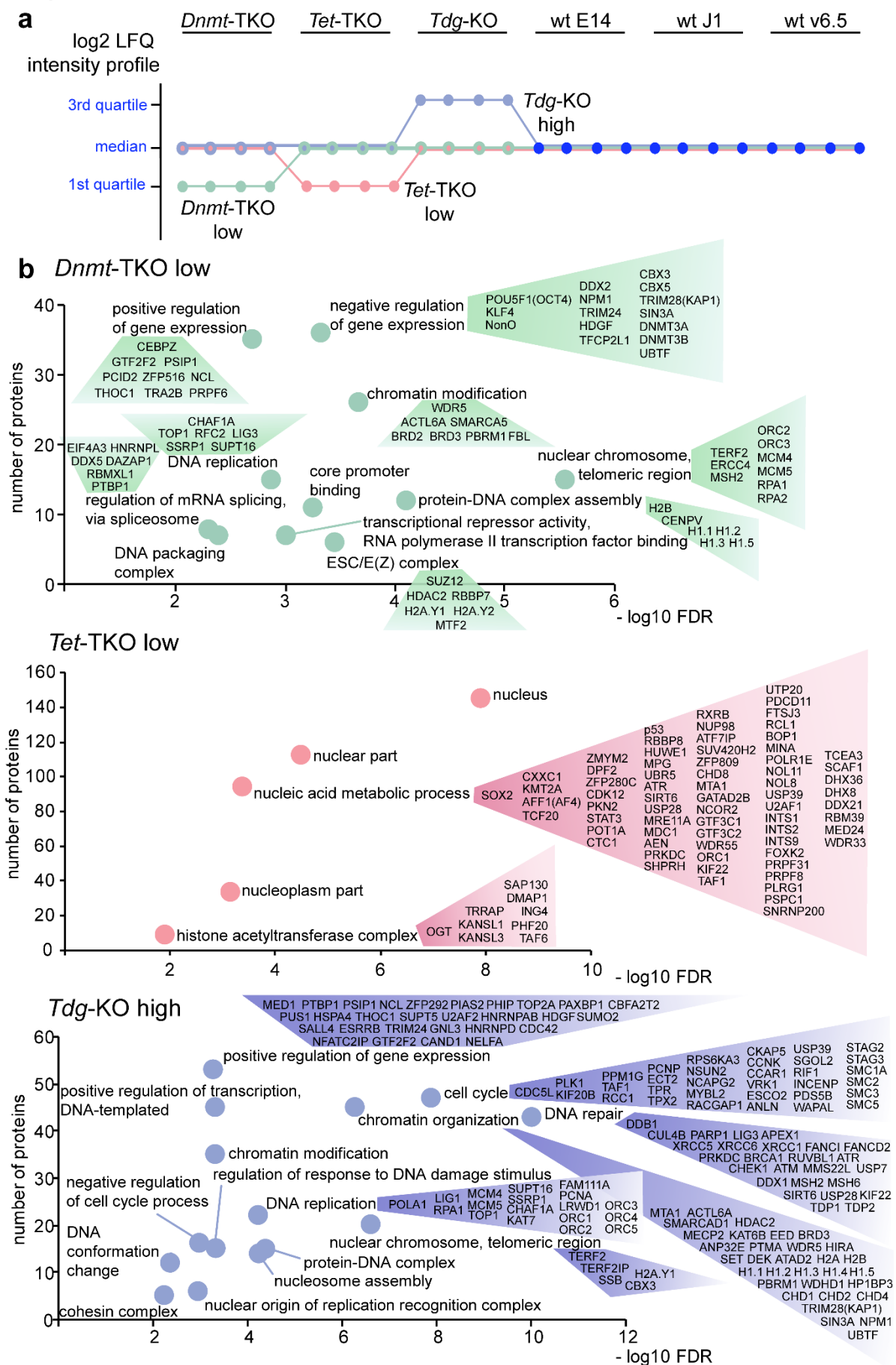
For analysis of modification-dependent chromatin composition, profile correlations were applied across samples separately for ESCs and EpiLCs. For ESCs, three different profile clusters were analysed: First, proteins that are depleted in the chromatin of *Dnmt*-TKOs (called "*Dnmt*-TKO low"), second, proteins under-represented in *Tet*-TKO chromatin ("*Tet*-TKO low"), and third, proteins highly abundant in chromatin of *Tdg*-TKOs ("*Tdg*-TKO high"). For this purpose, reference profiles have been created based on the first quartile, the third quartile, and the median of the dataset (Fig. 3a). Correlations were calculated selecting 25 neighbours and proteins with a profile FDR  $< 0.02$  have been defined as significant members of the cluster. GO-terms enriched in either of the three profile clusters are depicted in Figure 3b, raw LFQ profiles with highlighted examples are shown in Supplementary Figure 3. The three analysed clusters consist of 135 proteins significant for "*Dnmt*-TKO low", 183 proteins for "*Tet*-TKO low", and 234 proteins for "*Tdg*-TKO high".

The first cluster, "*Dnmt*-TKO low", contains proteins that are underrepresented in chromatin in the absence of functional DNA methyltransferases. Consequently, we find enrichment of GO-terms associated with transcription, interestingly both with positive and negative regulation of gene expression (FDR =  $2 \times 10^{-3}$  and FDR =  $5 \times 10^{-4}$ , respectively). As expected, all three DNMTs belong to this group, along with well-described heterochromatin proteins like SUZ12, HDAC1, HDAC2, CBX3, and CBX5. Additionally, we detect proteins of the telomeric region (FDR =  $3 \times 10^{-6}$ ), in agreement with a high DNA methylation density at telomere-proximal areas<sup>25</sup>. Finally, proteins associated with the GO-term "DNA packaging complex" (FDR =  $4 \times 10^{-3}$ ) include H2A.y (macro-H2A.1 and macro-H2A.2), and several isoforms of the linker histone H1 (H1.1, H1.2, H1.3, H1.5), highlighting the interplay between cytosine modifications and nucleosome organization.

The second cluster, "*Tet*-TKO low", represents proteins that do not bind to chromatin in the absence of TET dioxygenases. For this cluster, barely any specific GO-terms were enriched, except for "nucleus" and "nucleic acid metabolic process". TET1 and TET2 are detected in this group, confirming the robustness of the profile cluster correlations. Also, OGT is lowly abundant in *Tet*-TKO chromatin, since it is recruited by TET proteins<sup>26</sup>.

## 2. Results

**Figure 3**



**Figure 3: DNA modification-dependent chromatin composition in ESCs**

a) Reference profiles used for profile correlation analysis.  
 b) Selected GO-term enrichments for proteins lowly abundant in *Dnmt*-TKOs and *Tet*-TKOs and for proteins highly abundant in *Tdg*-TKOs as determined by profile correlation analysis. General GO-terms like "nucleus" or "DNA binding" have been removed in the *Dnmt*-TKO and *Tdg*-TKO plots for reasons of clarity. The plots depict the  $-\log_{10}$  FDR (Benjamini-Hochberg method) of the GO-term enrichment analysis on the x-axis and the number of proteins in the cluster that are associated with the corresponding GO-term on the y-axis. Selected proteins of the GO-groups are shown; candidates that belong to multiple GO-groups within one profile cluster are depicted only once. Supplementary Table 2 provides a comprehensive list of all proteins identified by profile correlation analysis, along with the associated FDR and the distance to the reference profile.

In the third cluster, "*Tdg*-KO high", we identify proteins that bind to chromatin in the presence of high levels of fC and caC. Confirming previous results<sup>18</sup>, "DNA repair" is one of the most significantly enriched GO-terms ( $\text{FDR} = 1 \times 10^{-10}$ ). Additionally, positive regulators of transcription and proteins associated with cell cycle progression are detected.

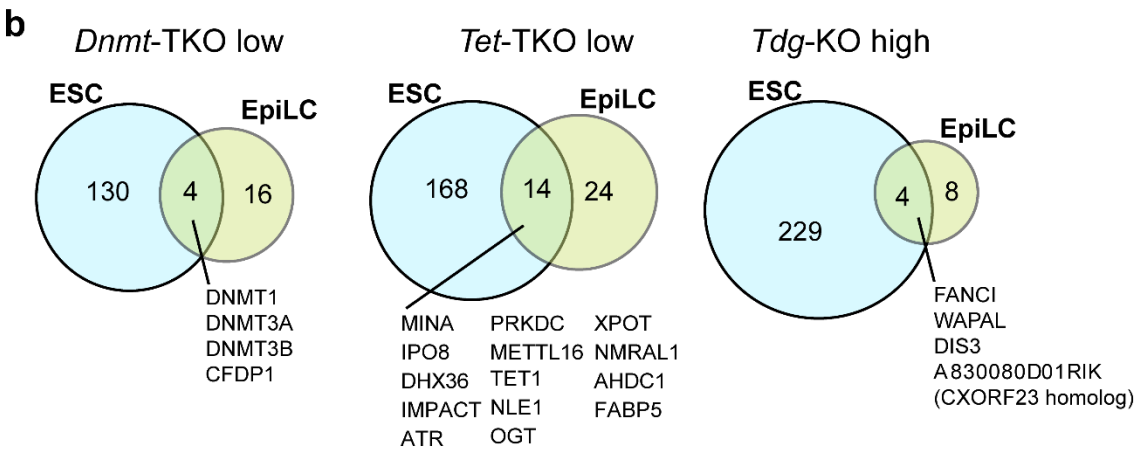
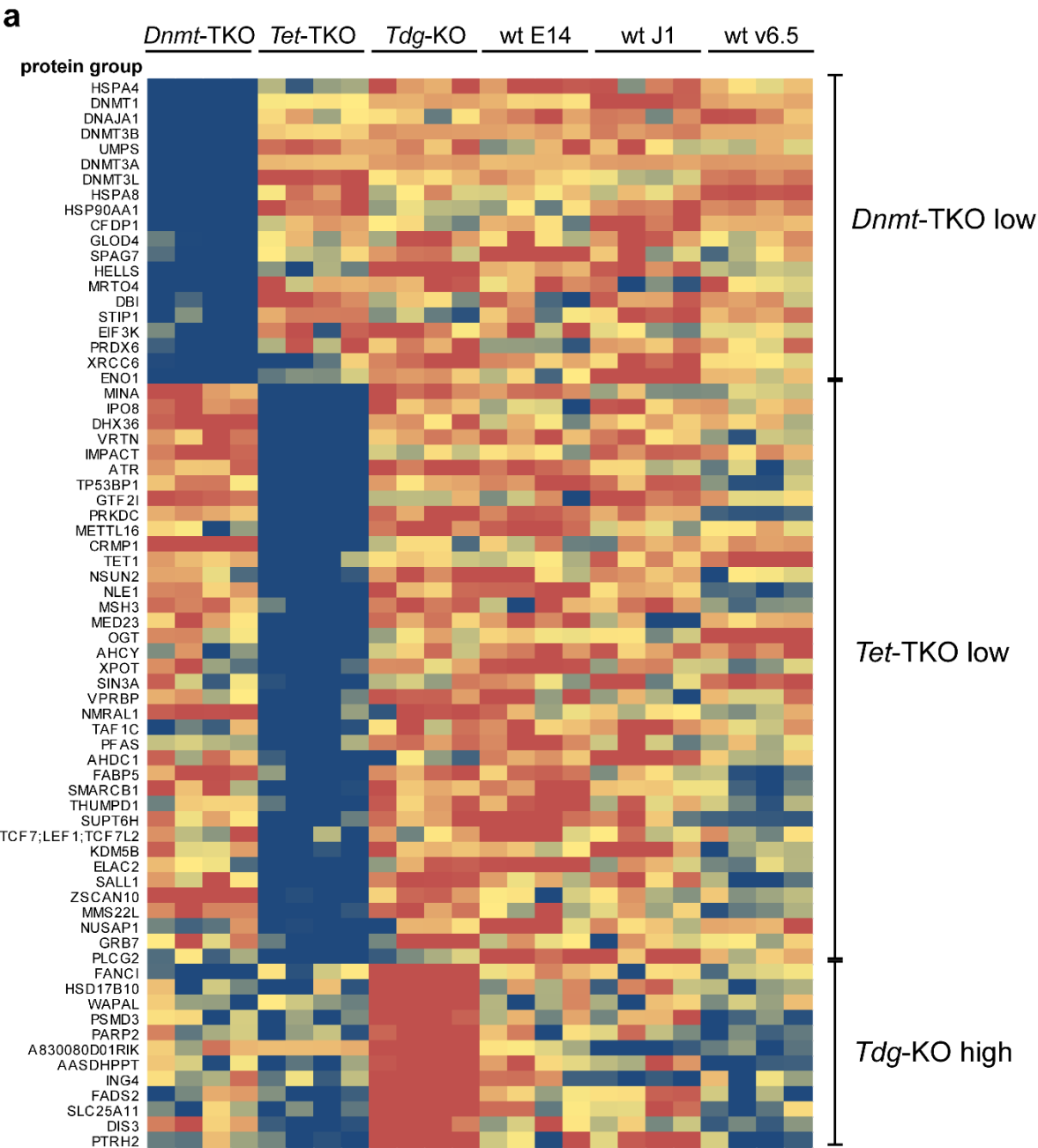
Interestingly, there is no overlap between proteins of the cluster "*Dnmt*-TKO low" and "*Tet*-TKO low". Contrarily, clusters "*Tdg*-KO high" and "*Dnmt*-TKO low" share 63 proteins and "*Tdg*-KO high" and "*Tet*-TKO low" share 26 proteins. E.g., the base excision repair proteins XRCC1 and LIG3 are lowly abundant in *Dnmt*-TKO chromatin, but highly present in *Tdg*-KO chromatin, suggesting a strong binding preference towards fC and/or caC.

**DNA modification-dependent chromatin in EpiLCs**

As with the analysis of ESC chromatin, profile correlation analysis was performed across all EpiLC ChEP samples with the clusters "*Dnmt*-TKO low", "*Tet*-TKO low", and "*Tdg*-KO high". In contrast to the ESC samples, the number of identified proteins per cluster was low, namely 21, 39, and 13 proteins, respectively. All significantly correlating proteins are shown as a heat map in Figure 4a. Comparison of the three analysed clusters in ESCs and EpiLCs shows that only a small fraction of proteins overlap (Fig. 4b). For *Dnmt*-TKO chromatin, all three DNMTs are lowly abundant as expected. Besides, CFDP1 is the only protein that loses its chromatin association both in *Dnmt*-TKO ESCs and *Dnmt*-TKO EpiLCs. CFDP1 is a poorly characterized protein that has been previously identified as a component of mitotic chromosomes in chicken and its yeast homologue Swc5 is a component of the SWR1 remodelling complex<sup>27, 28</sup>.

For the cluster "*Tet*-TKO low", 14 common proteins are diminished from chromatin in ESCs and EpiLCs. Among those are TET1 and OGT, and nuclear transporters like Importin-8 and Exportin-T. TET2 is only specifically absent from ESC chromatin in *Tet*-TKOs since its levels generally strongly decrease during EpiLC development (Fig. 2).

Figure 4



**Figure 4: Differential chromatin proteins in EpiLCs dependent on cytosine modifications**

a) Heat map of protein clusters in EpiLCs, which were identified by profile correlation analysis as depicted in Figure 3a. Log2 LFQ values have been normalized by Z-scoring prior to heat map generation; within clusters, proteins are sorted by ascending FDR (FDR cut-off 0.02).

b) Venn diagram of proteins of the clusters "*Dnmt*-TKO low", "*Tet*-TKO low", and "*Tdg*-KO high" in ESCs and EpiLCs. In the *Dnmt*-TKO cluster, 4 proteins are shared between the two cell states, in the *Tet*-TKO cluster 14 overlapping proteins are found, and in the *Tdg*-KO cluster, 4 proteins appear both in ESCs and EpiLCs.

Four proteins accumulate at chromatin in *Tdg*-KO cells both at the ESC and the EpiLC state: The repair protein FANCI, the negative regulator of chromatid cohesion WAPAL, the RNA degradation factor DIS3 and the uncharacterized protein CXORF23 homolog. Interestingly, PARP1 appears in the "*Tdg*-KO high" cluster in ESCs, whereas PARP2 occurs in EpiLCs.

**Discussion**

Although massive global genomic DNA methylation occurs during development from the ICM to the epiblast, *Dnmt*-KO cells can undergo this differentiation step<sup>9</sup>. *In vivo*, *Dnmt*-KO embryos develop normally until embryonic day 9.5 and die shortly after<sup>29,30</sup>. Consistent with these findings, we observe no morphological defects or increased cell death when differentiating *Dnmt*-TKO ESCs to EpiLCs. Consequently, we were able to compare the proteinaceous chromatin composition in ESCs and EpiLCs in three different wt cell lines (J1, v6.5, E14), *Dnmt*-TKOs, *Tet*-TKOs, and *Tdg*-KOs.

First, we examined the differences of wt ESC and EpiLC chromatin. Besides the well-described pluripotency factors KLF4, NR0B1, and ESRRB, we identify the transcriptional repressor TFCEP2L1 as a prominent determinant of ESC chromatin. In accordance with this result, TFCEP2L1 has been previously shown to be downregulated upon early primitive ectoderm differentiation<sup>31</sup>, a process that is induced by culturing ESCs in conditioned medium, leading to a cellular morphology similar to EpiLCs<sup>32</sup>. Also, microarray data of EpiLC differentiation suggest a strong downregulation of TFCEP2L1 expression as early as day 1 after induction<sup>19</sup>. Thus, the high abundance of TFCEP2L1 in ESC chromatin is likely due to transcriptional downregulation during EpiLC development.

Surprisingly, Epoxide hydrolase 2 (EPHX2, also sEH) is another protein strongly binding to ESC chromatin. This protein is supposed to mainly localize to the cytoplasm and to peroxisomes<sup>33</sup>. However, stainings of mouse brain endothelial cells show a weaker, but detectable nuclear signal in male cells<sup>34</sup>, suggesting that in certain cell types, EPHX2 might also be present in the nucleus. Similarly, ECHDC2, a very poorly characterized proteins, is highly present in ESC chromatin. Although a mitochondrial localization would be expected according to database annotations<sup>35</sup>, human ECHDC2 has been described to interact with DNA topoisomerase 3- $\alpha$ <sup>36</sup>, indicating a potential nuclear function. Since unspecific cross-linking in ChEP cannot be completely excluded,



## 2. Results

future experiments will have to show whether these two proteins have a specific role in ESC chromatin.

Another striking candidate significantly enriched in ESC chromatin as opposed to EpiLC chromatin is IFITM3 (also known as FRAGILIS). IFITM3 is reported to be expressed early in primordial germ cell (PGC) progenitors<sup>37</sup>. Thorough analysis of the mouse embryo at the gastrulation stage, however, revealed differential cellular localization patterns in different tissues: large cytoplasmic spots e.g. in epithelial cells and PGCs, cell surface localization e.g. in PGCs, and strong cytoplasmic and nuclear localization in visceral endoderm<sup>38</sup>. Having now found IFITM3 enriched in the chromatin of ESCs, this raises the interesting possibility that the protein is sequestered away from the nucleus during EpiLC differentiation.

Microarray data suggest a transcriptional downregulation upon development to EpiLCs for most of the proteins characteristic for ESC chromatin, except for TET2 and IFITM3 where coverage is rather low<sup>19</sup>. Recently, an ESC-specific enhancer of *Tet2* has been described that is silenced in EpiLCs<sup>39</sup>. Consistently, we have now described TET2 as a prominent determinant of chromatin of pluripotent ESCs that is depleted upon differentiation to EpiLCs.

As expected, chromatin of wt EpiLCs is predominantly characterized by a strong enrichment of DNMT3A and DNMT3B. Another protein highly enriched in EpiLC chromatin is the transcription factor SALL2 that binds to the consensus sequence GGG(C/T)GGG<sup>40</sup>. SALL2 has been shown to translocate into the nucleus upon treatment with nerve growth factor (NGF)<sup>41</sup> and might be responding to Activin A or FGF2, the two factors used to induce EpiLC differentiation, in a similar way.

Furthermore, the testis-specific linker histone variant H1t (HIST1H1T) is strongly present in EpiLC chromatin. Since PGCs are derived from epiblast cells, this finding suggests an early priming of EpiLCs towards the germ cell fate. The cell lines used in this study are all of male origin, and consequently, only the testis-specific histone variant H1t, but not the oocyte-specific variant H1oo, is detected in the ChEP dataset. The other more ubiquitous H1 isoforms H1.1 to H1.5 are not significantly enriched in EpiLC chromatin, but nevertheless show a certain drift towards EpiLCs (Fig. 2). This is in accordance with the fact that naïve pluripotent cells like ESCs have a more open and accessible genome organization<sup>42</sup>, and H1 is mostly found at higher order chromatin structures<sup>43</sup> established during differentiation.

The protein Cingulin (CGN), also highly present in EpiLC chromatin, is annotated as a tight junction protein<sup>35</sup>. However, stainings in the human cell lines IGROV 1 and Caco 2 show a nuclear localization in addition to localization at cellular borders<sup>44</sup>. Consistently, Cingulin has been suggested to also function as a transcription factor, regulating expression of tight junction proteins like Occludin or Claudins<sup>45</sup>. Since tight junction proteins have been discussed to influence gene expression and thereby integrate extracellular signals<sup>46</sup>, our finding that Cingulin specifically binds EpiLC chromatin further strengthens this point of view. Considering the drastic

morphological changes that ESCs undergo upon EpiLC differentiation, it is tempting to speculate that changes in cell adhesion trigger transcriptional changes and vice versa, possibly mediated by Cingulin.

After characterising wt ESC and EpiLC chromatin, we looked at proteins that change their abundance on the genome when major DNA modifiers like DNMTs, TETs, or TDG are depleted. In principle, there are three ways to explain these changes in abundance: (1) the detected proteins interact with the DNA modifier and are either recruited or repelled by its presence; (2) the proteins intrinsically prefer or disfavour to bind a certain cytosine variant generated by the DNA modifier; and (3) the expression level of the protein is altered due to epigenetic changes induced by KO of *Dnmt1/3a/3b*, *Tet1/2/3*, or *Tdg*.

One of the proteins most significantly depleted from chromatin in *Dnmt*-TKO ESCs is the transcription and splicing regulator NonO, which has been shown to bind to intracisternal-A particles (IAP) proximal enhancer elements<sup>47</sup>. Since IAP sequences are heavily demethylated upon loss of all three DNMTs<sup>20</sup>, this finding suggests that NonO requires DNA methylation in order to bind to its target.

Additionally, we find the two Yamanaka factors<sup>48</sup> KLF4 and POU5F1 (better known as OCT4) depleted from chromatin in *Dnmt*-TKO ESCs. In line with this observation, KLF4 has recently been described to preferentially bind to its consensus sequence, GGGCGTG, when the C is methylated<sup>18</sup>. Collectively, these results suggest that OCT4 and KLF4 need a minimum mC density in order to bind to chromatin efficiently. Interestingly, a third Yamanaka factor, SOX2, is found in the "*Tet*-TKO low" cluster, implying either a preference for hmC, fC, or caC, or a repulsion by DNA hypermethylation. Taken together, our data provide the first evidence that the DNA modification status may be critical for maintaining the presence of KLF4, OCT4, and SOX2 on chromatin.

Interestingly, CXXC1 is also specifically diminished from *Tet*-TKO chromatin in ESCs. CXXC1 localizes to unmethylated CpG motifs, and its KO leads to global DNA hypomethylation as well as differentiation defects in very early embryonic development around the implantation stage<sup>49, 50</sup>. Since CXXC1 binding to chromatin mostly depends on protein-protein interactions and not on direct DNA binding affinity<sup>51</sup>, our data suggest that CXXC1 and TET proteins might have a cooperative role in chromatin regulation. Furthermore, OGT, which co-localizes with TET1 at CpG-rich unmethylated transcription start sites<sup>52</sup>, has recently been found to interact with CXXC1<sup>53</sup>. Since both OGT and CXXC1 dissociate from chromatin in *Tet*-TKO ESCs, we suggest a functional link between TET1, OGT, and CXXC1 in transcriptionally active CpG-dense genomic regions.

Chromatin of *Tdg*-KO ESCs is most significantly characterized by a high abundance of DNA repair factors. Besides proteins involved in base excision repair (PARP1, LIG3, APEX1, XRCC1), previously reported to play a role in removal of fC and caC<sup>3, 4, 23</sup>, we also find central components of the mismatch repair (MSH2, MSH6, HMGB1) and non-homologous end-joining pathways (PRKDC, XRCC5, XRCC6). In addition to repair factors, proteins with connections to cell cycle and

## 2. Results

DNA replication are highly enriched. These aspects might be linked, since G-caC base pairing has been shown to trigger mismatch repair and affect cell proliferation<sup>54</sup>. Interestingly, we specifically identify the components of the MutS $\alpha$  heterodimer (MSH2, MSH6) associated with *Tdg*-KO chromatin, a complex that recognizes mostly base-base mismatches<sup>55, 56</sup>. Mismatch repair is coupled to DNA replication<sup>56</sup>, and consequently, we find proteins such as PCNA and LIG1 enriched. Furthermore, proteins involved in replication initiation like ORCs or MCMs are detected, as well as the catalytic subunit of DNA polymerase alpha (POLA1).

Intriguingly, several components of sister chromatid cohesion complexes (PDS5B, WAPAL, STAG and SMC proteins) strongly associate with *Tdg*-KO chromatin. Again, this might be due to an increased need for DNA repair, since cohesion complexes are especially important in double strand break repair<sup>57</sup>.

In contrast to chromatin in ESCs, we find only few proteins with differential abundances in EpiLCs deficient for DNA modifiers. This might be largely due to statistical effects since the EpiLC dataset is more heterogeneous also among wt cell lines and thus, there is more background fluctuation of protein abundances. Nevertheless, the correlation analysis seems to be robust, because the knocked-out proteins are detected in the "*Dnmt*-TKO low" or "*Tet*-TKO low" clusters serving as a proof-of-principle. Additionally, DNMT3L, a known interactor and important co-factor of the *de novo* methyltransferases<sup>58</sup>, is specifically depleted from *Dnmt*-TKO chromatin in EpiLCs.

Interestingly, the protein HELLS (also known as LSH) shows significantly lower levels in EpiLC chromatin in the absence of DNMTs. KO of *Hells* leads to severe global DNA hypomethylation<sup>59</sup>, and the protein has been suggested as a recruiting factor for DNMT1 and DNMT3B<sup>60</sup>. Our data imply that HELLS requires either the presence of DNMT proteins or mC to efficiently bind to the genome. Consistent with this observation, HELLS has been shown to depend on intact heterochromatin organization for chromatin binding, and to lose its association with pericentric heterochromatin upon treatment with histone deacetylase inhibitors<sup>61</sup>. KO of DNMTs could have a similar effect by perturbing proper chromatin condensation and thus leading to mislocalization of HELLS.

In *Tet*-TKO EpiLCs, VprBP and SIN3A show low chromatin abundance, two proteins previously shown to interact with TET proteins<sup>62, 63</sup>. Furthermore, SALL1, a transcription factor involved in regulation of pluripotency via association with NANOG<sup>64</sup>, is also depleted from chromatin. Similarly, TET1 interacts with NANOG and binds loci regulating pluripotency and differentiation<sup>65</sup>. Our finding suggests an interplay between TET1 and SALL1 in EpiLCs, possibly regulating exit from ground state pluripotency. The chromatin of EpiLCs deficient for *Tdg* shows very little changes. This might be explained by the fact that accumulation of fC and caC is far less pronounced in *Tdg*-KO EpiLCs when compared to *Tdg*-KO ESCs (Fig. 1b), supporting an essential role for the TET-TDG demethylation pathway in pluripotency.

In summary, we provide a comprehensive analysis of chromatin proteins in ESCs and EpiLCs. Furthermore, we investigated the chromatin composition if major DNA modifying enzymes like DNMTs, TETs, or TDG are depleted. Our data not only serve as a resource to understand the interplay between DNA modifications and proteinaceous chromatin, but also contribute to comprehend the transition from naïve pluripotent ESCs to the primed cells of the epiblast.

## Methods

### Cell culture

ESCs were cultured and differentiated as described previously<sup>66</sup>. Transition of ESCs to EpiLCs was induced based on the protocol by Hayashi *et al*<sup>19</sup>. Briefly, cells were kept in N2B27/2i/LIF medium under serum-free and feeder-free conditions for a minimum of 7 days prior to induction of EpiLC differentiation. 12 ng/ml Fgf2 (Peprotech), 20 ng/ml Activin A (Peprotech) and 1 % Knockout Serum Replacement (Life Technologies) in N2B27 were used to start EpiLC differentiation at a seeding density of  $4.6 \times 10^4$  cells/cm<sup>2</sup>. EpiLCs were harvested 64 h after induction of differentiation. Per replicate, one T175 flask of EpiLCs and 2 T175 flasks of ESCs (approximately  $50 \times 10^6$  cells) were trypsinised, pelleted at 500 g for 2 min, washed with PBS (Sigma) and immediately used for ChEP (Chromatin Enrichment for Proteomics).

### Microscopy

Phase contrast images were acquired with an EVOS FL Cell Imaging System (Life Technologies) using a UPLAN 4x objective (NA 0.13). For high-throughput imaging, cells were grown in 96-well microplates (μClear, Greiner Bio-One), washed with PBS and fixed with 4 % formaldehyde in PBS for 10 min at room temperature (RT). The fixative was gradually exchanged to PBST (0.02 % Tween/PBS) and cells were washed twice with PBST. Cells were permeabilised (0.5 % Triton-X100/PBST) for 10 min at RT, washed twice with PBST, and then treated with denaturation solution (2 N HCl) for 45 min at RT. After washing with PBST, cells were incubated with renaturation solution (150 mM Tris-HCl, pH 8.5) for 30 min at RT and washed twice with PBST. Cells were blocked (2% BSA/PBST) for 45 min at RT and incubated with primary antibodies (mouse-anti-mC, Diagenode 33D3; rabbit-anti-hmC, active-motif 39769; rabbit-anti-fC, active-motif 61223; rabbit-anti-caC, active-motif 61225) for 1 h at 37 °C. After washing three times with PBST, cells were incubated with secondary antibodies (goat-anti-mouse or goat-anti-rabbit coupled to Alexa594, Thermo Fisher) for 1 h at 37 °C. Cells were washed three times with PBST, counterstained with 200 ng/ml DAPI, and covered with PBS. Images were acquired with the Operetta high-content image analysis system (PerkinElmer, 40x high NA objective) followed by analysis with the Harmony software (PerkinElmer). DAPI was used for the detection of single nuclei and cytosine modifications were quantified in the selected nuclei based on the antibody

## 2. Results

signal intensity. Per cell line and modification, 1500 to 7500 nuclei were used for quantification. Boxplots and statistical analysis were carried out with GraphPad Prism 6.

### **Chromatin enrichment for proteomics (ChEP)**

Chromatin enrichment was based on the protocol by Kustatscher *et al*<sup>24,67</sup>. Cell pellets were cross-linked with 200 µl 1 % formaldehyde (Polysciences) for 10 min at 37 °C and the reaction was subsequently quenched with 0.25 M glycine for 5 min at RT. After washing with 500 µl PBS (Sigma) and centrifugation (500 g, 5 min), cells were lysed in 200 µl ice-cold lysis buffer (25 mM TrisHCl, pH=7.4, 0.1 % Triton, 85 mM KCl, protease inhibitor mix (Roche)). Nuclei were pelleted at 2300 g for 5 min at 4 °C, resuspended in 100 µl lysis buffer (containing 200 µg/ml RNase A (Applichem)) and incubated for 15 min at RT. After centrifugation (2300 g, 10 min, 4 °C), nuclei were resuspended in 100 µl SDS buffer (50 mM TrisHCl, pH=7.4, 10 mM EDTA, 4 % SDS, protease inhibitor mix (Roche)), incubated for 10 min at RT, and 300 µl of urea buffer (10 mM TrisHCl, pH=7.4, 1 mM EDTA, 8 M urea) were added. Chromatin was precipitated at 21,000 g for 30 min at 15 °C and once more resuspended in 100 µl SDS buffer, mixed with 300 µl urea buffer, and pelleted at 21,000 g for 30 min at 15 °C. The resulting chromatin fraction (gel-like pellet) was covered with 100 µl storage buffer (10 mM TrisHCl, pH=7.4, 1 mM EDTA, 25 mM NaCl, 10 % glycerol, protease inhibitor mix (Roche)) and sonicated with a Bioruptor (Diagenode) set to "High" for 15 min (30 s pulse, 60 s pause). Insoluble chromatin was precipitated at 21,000 g for 30 min at 4 °C, the supernatant was quantified, incubated with 50 µl PSB-TCEP (Macherey-Nagel) for 30 min at 95 °C, and stored at -20 °C. Prior to processing for mass spectrometry, samples were incubated for 30 min at 95 °C and run over a 10 % SDS gel (Biorad) at 50 V for approximately 30 min.

### **In-gel digestion**

Each lane was treated as two fractions, one containing the thick lower molecular weight band and the other containing all other proteins. The gel pieces were cut approximately into 1 mm x 1 mm size and subjected to standard in-gel digestion. Afterwards, the peptides were extracted and concentrated on a SDB-XC StageTip<sup>68</sup>. The StageTips were washed twice with 0.1 % formic acid before eluting. Eluted samples were dried completely in a SpeedVac concentrator and resuspended in 6 µl of 0.1 % formic acid of which 5 µl were loaded onto the column with the autosampler.

### **LC-MS/MS analysis**

The purified peptides were loaded onto a 15 cm long , 75 µm inner diameter reversed phase column (New Objective), packed with 1.9 µm C18 particles (Dr. Maisch GmbH) using the autosampler Thermo Easy nLC 1000. The column was maintained at a constant temperature of 45 °C in order to improve the peptide retention time reproducibility and the resolution of the separation. Peptides were separated over a 2 h gradient and directly sprayed into the orifice of a

Q Exactive HF mass spectrometer (Thermo Scientific). The mass spectrometry data were generated using data dependent acquisition methods. Survey scans were recorded at a resolution of 120,000 (at  $m/z = 200$ ) and up to top 10 most intense precursors were subjected to MS/MS events with a precursor intensity threshold of  $10^5$  charges and all MS/MS scans were recorded at a resolution of 15,000 ( $m/z = 200$ ). In order to minimize repeat sequencing, dynamic exclusion was set for 30 s, only 2+, 3+, 4+, and 5+ charged precursors were selected for fragmentation. All data were recorded in profile mode.

### Proteomics data analysis and statistics

MaxQuant<sup>69</sup> version 1.5.2.8 was used to analyse Thermo Scientific raw files, determine peptide sequences, assign peptides to protein groups (using the reviewed ("Swiss-Prot") UniprotKB mouse proteome as reference) and quantify protein groups with the label-free quantification (LFQ) algorithm<sup>70</sup>. The "match-between-runs" option was applied, allowing for a maximum of 3 missed cleavages and treating protein N-terminal acetylation and methionine oxidation as variable modifications. Carbamidomethylation of cysteine was set as fixed modification. For histone proteins (Uniprot-IDs P43275, P43276, P15864, P43277, P43274, P22752, P70696, P10853, P10854, P68433, P84228, P62806), the following variable modifications were used: lysine acetylation, lysine/arginine dimethylation, lysine ubiquitination (K-GlyGly motif), hydroxyproline, lysine/arginine monomethylation, OHexNAc, serine/threonine/tyrosine phosphorylation, and lysine trimethylation.

Downstream data analysis and statistics was performed with the MaxQuant-associated Perseus software version 1.5.0.15. The complete dataset consisting of 5729 protein groups was filtered prior to analysis in the following way: After removal of common contaminants like keratins, the minimal number of unique peptides per protein group was set to 2, requiring that protein group to be quantifiable in at least 3 out of 4 replicates in at least one set of replicates. Missing values were imputed by a normal distribution of lower values with a downshift of 1.8 and a width of 0.3 relative to the standard deviation of the sample. Two-way ANOVA with a cut-off of  $p < 0.05$  ( $-\log p > 1.3$ ) for either of the two variables or an interaction thereof was used to further filter the data. For the ANOVA, the following variables were defined depending on the analysis: For evaluation of the whole dataset (Fig. 1c), variable (1) was set as ESC or EpiLC state, and variable (2) as wt, *Dnmt*-TKO, *Tet*-TKO, or *Tdg*-KO cell line. For analysis of wt chromatin (Fig. 2), variable (1) was set as ESC or EpiLC state, and variable (2) as J1, v6.5, or E14. For separate analysis of ESC and EpiLC chromatin (Fig. 3 and 4), variable (1) was set as KO, J1, v6.5, or E14 cell line, and variable (2) as wt, *Dnmt*-TKO, *Tet*-TKO, or *Tdg*-KO cell line. Profile correlations were performed with Perseus 1.5.0.15, calculating profile FDRs based on the selection of 25 neighbours. Reference profiles were constructed based on the 1st quartile border ( $\log_2$  LFQ intensity = 26.62 for ESCs, 27.28 for EpiLCs) for "low" values, 3rd quartile border ( $\log_2$  LFQ intensity = 29.80 for ESCs, 30.27 for



## 2. Results

EpiLCs) for "high" values, with all other values set to the median (log2 LFQ intensity = 28.16 for ESCs, 28.74 for EpiLCs). GO-term enrichment analysis was performed with the online tool GOrilla<sup>71</sup>.

## Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) (SFB 646/B10 and SFB 1064/A17 to HL). CB and CT are supported by the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS). HL is a member of the Nanosystems Initiative Munich (NIM). We are grateful to the following researchers for providing cell lines: R. Jaenisch and F. Lyko for *Tet*-TKO and v6.5, M. Okano for *Dnmt*-TKO, and P. Schär for *Tdg*-KO.

## References

1. Bestor, T.H. The DNA methyltransferases of mammals. *Hum Mol Genet* 9, 2395-2402 (2000).
2. Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K. & Yura, R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem J* 126, 781-790 (1972).
3. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303-1307 (2011).
4. Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286, 35334-35338 (2011).
5. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930-935 (2009).
6. Auclair, G. & Weber, M. Mechanisms of DNA methylation and demethylation in mammals. *Biochimie* 94, 2202-2211 (2012).
7. Iqbal, K., Jin, S.G., Pfeifer, G.P. & Szabo, P.E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A* 108, 3642-3647 (2011).
8. Inoue, A., Shen, L., Matoba, S. & Zhang, Y. Haploinsufficiency, but not defective paternal 5mC oxidation, accounts for the developmental defects of maternal Tet3 knockouts. *Cell Rep* 10, 463-470 (2015).
9. Auclair, G., Guibert, S., Bender, A. & Weber, M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol* 15, 545 (2014).
10. Shen, L. *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 153, 692-706 (2013).
11. Wu, H. *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev* 25, 679-684 (2011).
12. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem* 6, 1049-1055 (2014).
13. Otani, J. *et al.* Cell Cycle-Dependent Turnover of 5-Hydroxymethyl Cytosine in Mouse Embryonic Stem Cells. *PLoS One* 8, e82961 (2013).
14. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat Chem Biol* 11, 555-557 (2015).
15. Fatemi, M. & Wade, P.A. MBD family proteins: reading the epigenetic code. *J Cell Sci* 119, 3033-3037 (2006).
16. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* 18, 6538-6547 (1998).
17. Nan, X., Meehan, R.R. & Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res* 21, 4886-4892 (1993).

18. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146-1159 (2013).
19. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* 146, 519-532 (2011).
20. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* 11, 805-814 (2006).
21. Dawlaty, M.M. *et al.* Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell* 29, 102-111 (2014).
22. Cortazar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* 470, 419-423 (2011).
23. Muller, U., Bauer, C., Siegl, M., Rottach, A. & Leonhardt, H. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res* 42, 8592-8604 (2014).
24. Kustatscher, G. *et al.* Proteomics of a fuzzy organelle: interphase chromatin. *EMBO J* 33, 648-664 (2014).
25. Brock, G.J., Charlton, J. & Bird, A. Densely methylated sequences that are preferentially localized at telomere-proximal regions of human chromosomes. *Gene* 240, 269-277 (1999).
26. Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 493, 561-564 (2012).
27. Ohta, S. *et al.* The Protein Composition of Mitotic Chromosomes Determined Using Multiclassifier Combinatorial Proteomics. *Cell* 142, 810-821 (2010).
28. Iwashita, S. *et al.* Mammalian Bcnt/Cfdp1, a potential epigenetic factor characterized by an acidic stretch in the disordered N-terminal and Ser250 phosphorylation in the conserved C-terminal regions. *Bioscience Reports* 35, e00228-e00228 (2015).
29. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915-926 (1992).
30. Okano, M., Bell, D.W., Haber, D.A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247-257 (1999).
31. Rodda, S., Sharma, S., Scherer, M., Chapman, G. & Rathjen, P. CRTR-1, a developmentally regulated transcriptional repressor related to the CP2 family of transcription factors. *J Biol Chem* 276, 3324-3332 (2001).
32. Rathjen, J. *et al.* Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *J Cell Sci* 112 ( Pt 5), 601-612 (1999).
33. Fransen, M., Nordgren, M., Wang, B. & Apanasets, O. Role of peroxisomes in ROS/RNS-metabolism: implications for human disease. *Biochim Biophys Acta* 1822, 1363-1373 (2012).
34. Gupta, N.C., Davis, C.M., Nelson, J.W., Young, J.M. & Alkayed, N.J. Soluble epoxide hydrolase: sex differences and role in endothelial cell survival. *Arterioscler Thromb Vasc Biol* 32, 1936-1942 (2012).
35. UniProt, C. UniProt: a hub for protein information. *Nucleic Acids Res* 43, D204-212 (2015).
36. Huttlin, E.L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 162, 425-440 (2015).
37. Saitou, M., Barton, S.C. & Surani, M.A. A molecular programme for the specification of germ cell fate in mice. *Nature* 418, 293-300 (2002).
38. Mikedis, M.M. & Downs, K.M. Widespread but tissue-specific patterns of interferon-induced transmembrane protein 3 (IFITM3, FRAGILIS, MIL-1) in the mouse gastrula. *Gene Expr Patterns* 13, 225-239 (2013).
39. Sohni, A. *et al.* Dynamic switching of active promoter and enhancer domains regulates Tet1 and Tet2 expression during cell state transitions between pluripotency and differentiation. *Mol Cell Biol* 35, 1026-1042 (2015).
40. Gu, H. *et al.* DNA-binding and regulatory properties of the transcription factor and putative tumor suppressor p150(Sal2). *Biochim Biophys Acta* 1809, 276-283 (2011).
41. Pincheira, R., Baerwald, M., Dunbar, J.D. & Donner, D.B. Sall2 is a novel p75NTR-interacting protein that links NGF signalling to cell cycle progression and neurite outgrowth. *EMBO J* 28, 261-273 (2009).
42. Ricci, M.A., Manzo, C., Garcia-Parajo, M.F., Lakadamyali, M. & Cosma, M.P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* 160, 1145-1158 (2015).

## 2. Results

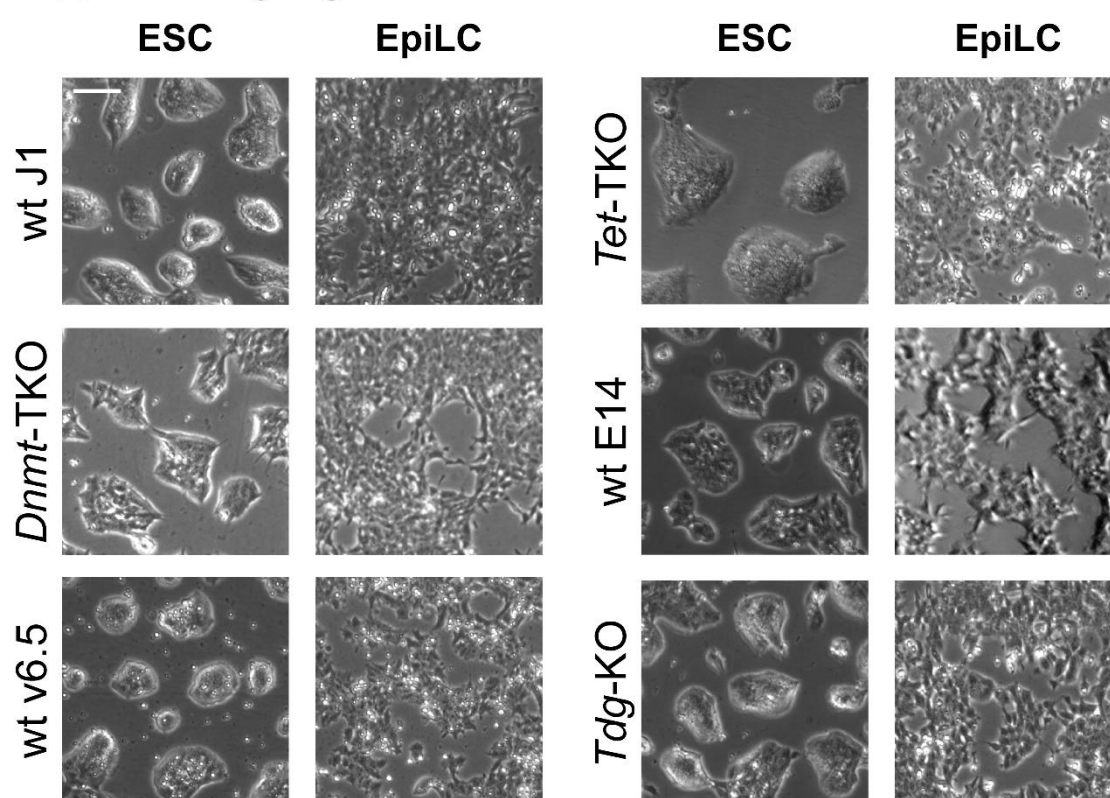
43. Happel, N. & Doenecke, D. Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* 431, 1-12 (2009).
44. Nakamura, T. *et al.* huASH1 protein, a putative transcription factor encoded by a human homologue of the *Drosophila ash1* gene, localizes to both nuclei and cell-cell tight junctions. *Proc Natl Acad Sci U S A* 97, 7284-7289 (2000).
45. Guillemot, L. *et al.* Disruption of the cingulin gene does not prevent tight junction formation but alters gene expression. *J Cell Sci* 117, 5245-5256 (2004).
46. Balda, M.S. & Matter, K. Tight junctions and the regulation of gene expression. *Biochim Biophys Acta* 1788, 761-767 (2009).
47. Basu, A., Dong, B., Krainer, A.R. & Howe, C.C. The intracisternal A-particle proximal enhancer-binding protein activates transcription and is identical to the RNA- and DNA-binding protein p54nrb/NonO. *Mol Cell Biol* 17, 677-686 (1997).
48. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663-676 (2006).
49. Carlone, D.L. *et al.* Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. *Mol Cell Biol* 25, 4881-4891 (2005).
50. Carlone, D.L. & Skalnik, D.G. CpG binding protein is crucial for early embryonic development. *Mol Cell Biol* 21, 7601-7606 (2001).
51. Lee, J.H. & Skalnik, D.G. CpG-binding protein is a nuclear matrix- and euchromatin-associated protein localized to nuclear speckles containing human trithorax. Identification of nuclear matrix targeting signals. *J Biol Chem* 277, 42259-42267 (2002).
52. Vella, P. *et al.* Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol Cell* 49, 645-656 (2013).
53. Hein, M.Y. *et al.* A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163, 712-723 (2015).
54. Shibutani, T. *et al.* Guanine- 5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Sci Rep* 4, 5220 (2014).
55. Drummond, J.T., Li, G.M., Longley, M.J. & Modrich, P. Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* 268, 1909-1912 (1995).
56. Li, G.M. Mechanisms and functions of DNA mismatch repair. *Cell Res* 18, 85-98 (2008).
57. Peters, J.M., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev* 22, 3089-3114 (2008).
58. Gowher, H., Liebert, K., Hermann, A., Xu, G. & Jeltsch, A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J Biol Chem* 280, 13341-13348 (2005).
59. Dennis, K., Fan, T., Geiman, T., Yan, Q. & Muegge, K. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev* 15, 2940-2944 (2001).
60. Myant, K. & Stancheva, I. LSH cooperates with DNA methyltransferases to repress transcription. *Mol Cell Biol* 28, 215-226 (2008).
61. Yan, Q., Cho, E., Lockett, S. & Muegge, K. Association of Lsh, a Regulator of DNA Methylation, with Pericentromeric Heterochromatin Is Dependent on Intact Heterochromatin. *Mol Cell Biol* 23, 8416-8428 (2003).
62. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343-348 (2011).
63. Nakagawa, T. *et al.* CRL4(VprBP) E3 ligase promotes monoubiquitylation and chromatin binding of TET dioxygenases. *Mol Cell* 57, 247-260 (2015).
64. Karantzali, E. *et al.* Sall1 regulates embryonic stem cell differentiation in association with nanog. *J Biol Chem* 286, 1037-1045 (2011).
65. Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* 495, 370-374 (2013).
66. Mulholland, C.B. *et al.* A modular open platform for systematic functional studies under physiological conditions. *Nucleic Acids Res* 43, e112 (2015).
67. Kustatscher, G., Wills, K.L., Furlan, C. & Rappsilber, J. Chromatin enrichment for proteomics. *Nat Protoc* 9, 2090-2099 (2014).
68. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 75, 663-670 (2003).

69. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26, 1367-1372 (2008).
70. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13, 2513-2526 (2014).
71. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48 (2009).

## Supplementary Material

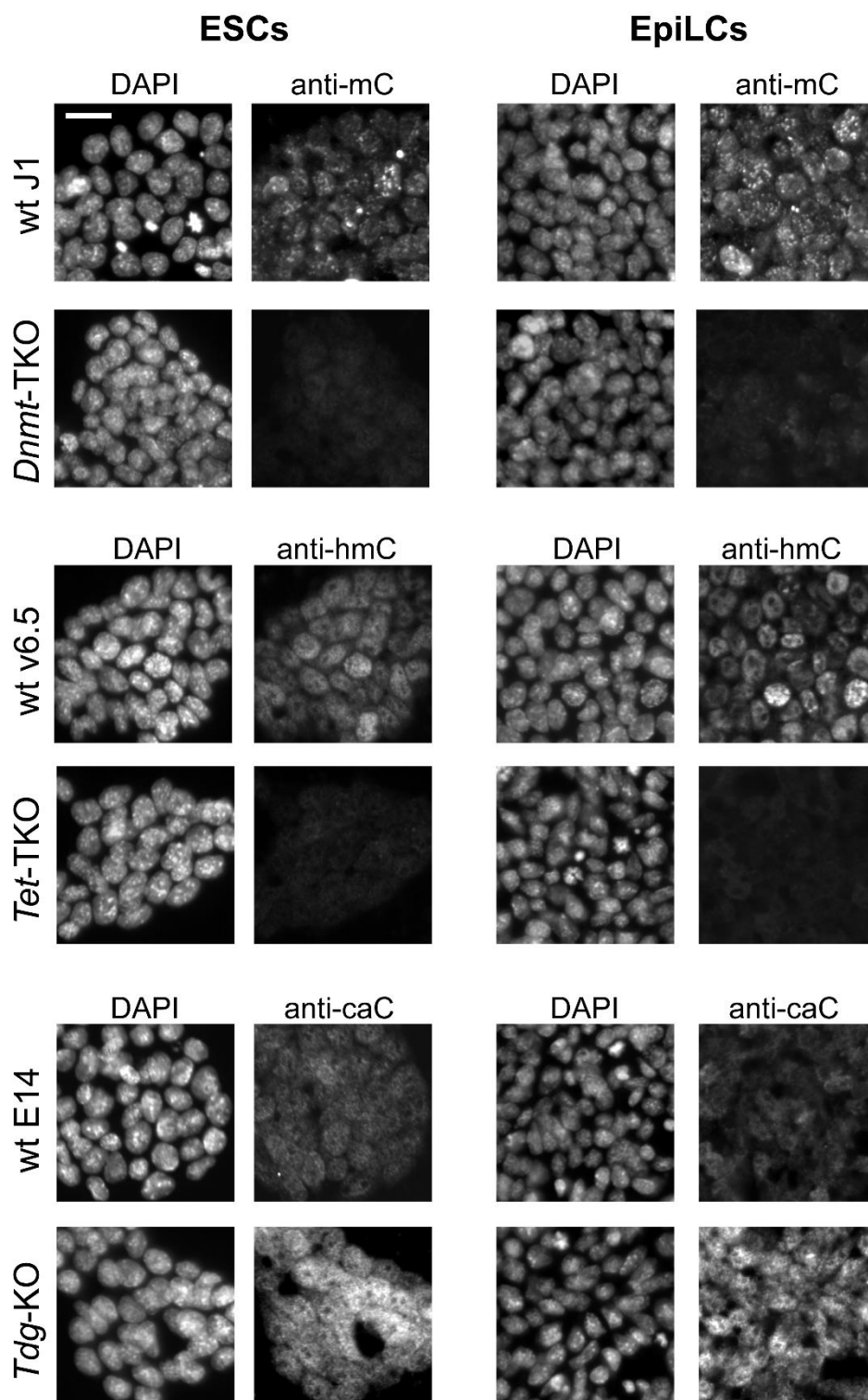
### Supplementary Figures

#### Supplementary Figure 1



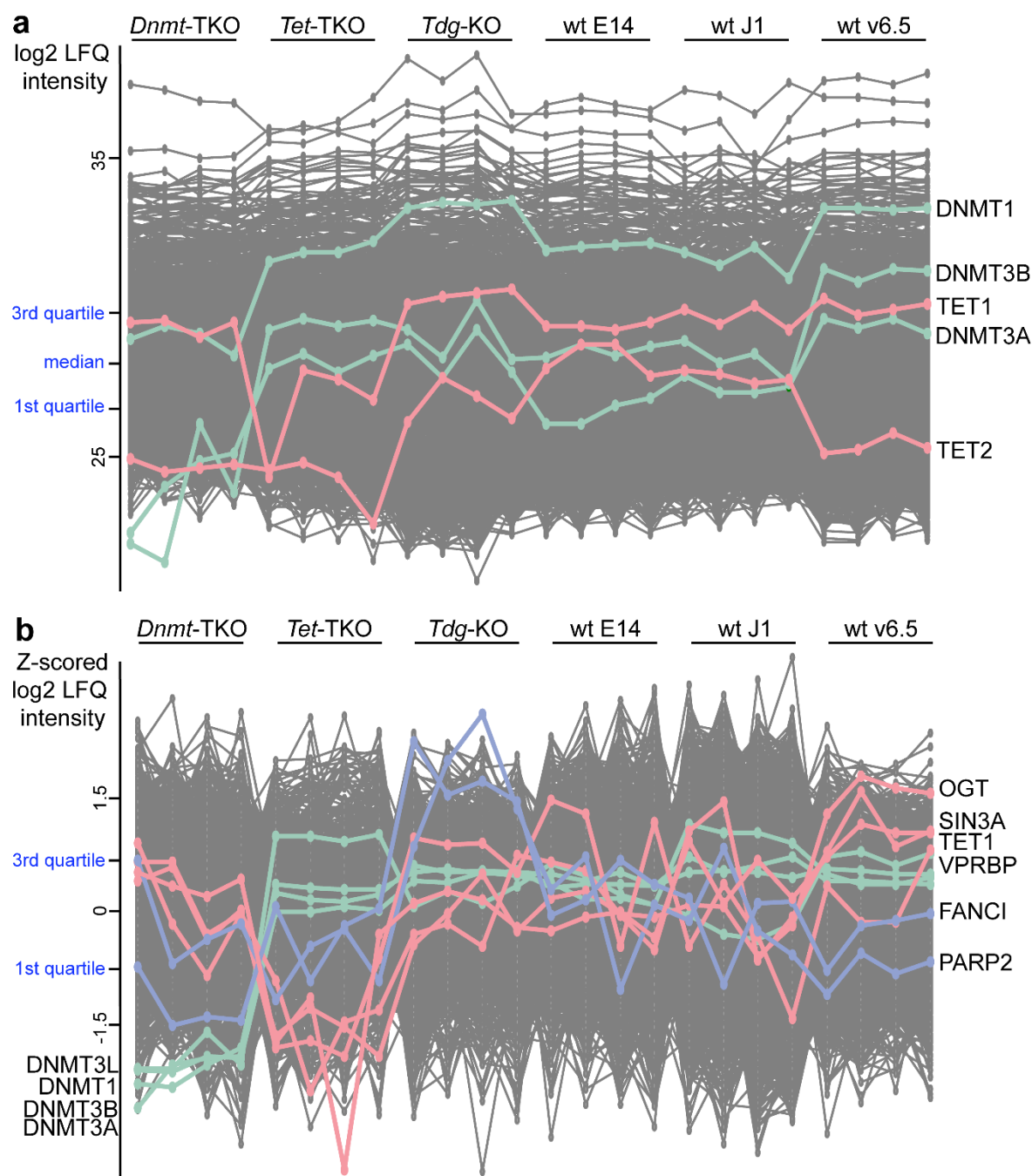
#### Supplementary Figure 1: Morphology of ESCs and EpiLCs

Phase contrast images of all six cell lines in the ESC state and after 40 h of EpiLC differentiation. scale bar: 100  $\mu$ m

**Supplementary Figure 2****Supplementary Figure 2: Immunofluorescence of cytosine modifications**

Examples of images acquired with the Operetta high-content image analysis system that were used for quantification of cytosine modifications. Constant acquisition settings were applied across all samples. Scale bar: 10  $\mu$ m

### Supplementary Figure 3



#### Supplementary Figure 3: LFQ profiles in ESCs and EpiLCs

a) Log<sub>2</sub> LFQ intensity profile plot across all ChEP samples in ESCs. Profiles for DNMTs and TETs are highlighted in green and red, respectively. The median and the border for the first and third quartile are indicated in blue.

b) Profile plot of all proteins identified in EpiLCs. For visualization purposes, log<sub>2</sub> LFQ intensities have been Z-scored before plotting. The borders for the first and third quartile are indicated in blue. Examples of proteins of the clusters “*Dnmt*-TKO low”, “*Tet*-TKO low”, and “*Tdg*-KO high” are highlighted in green, red, and light blue, respectively.

## 2. Results

### Supplementary Tables

#### **Supplementary Table 1: Significantly different proteins in EpiLCs versus ESCs identified by ChEP**

Perseus export of wt EpiLC versus wt ESC volcano plot. Proteins with negative  $\Delta \log_2$  LFQ are significantly enriched in the ESC sample, proteins with positive  $\Delta \log_2$  LFQ in the EpiLC sample.

<b>-log p</b>	<b><math>\Delta \log_2</math> LFQ</b>	<b>Gene names</b>			
			5.66	-1.97	Mlec
			4.71	-1.96	Gss
15.17	-5.07	Klf4	3.45	-1.96	Hint2
10.97	-5.03	Tfcp2l1	3.58	-1.93	Praf2
9.62	-4.39	Esrrb	5.76	-1.92	Bcat1
12.88	-3.78	Ephx2	1.64	-1.92	Rpl39
7.15	-3.53	Tet2	4.29	-1.92	Arih2
6.52	-3.40	Slc2a3	4.22	-1.89	E130012A19Rik
3.99	-3.31	Rev3l	3.81	-1.87	Oat
8.12	-3.16	Arid5b	3.56	-1.87	Cd9
6.17	-3.10	Fabp3	8.99	-1.86	Zswim3
7.99	-3.08	Nr0b1	1.90	-1.86	Tomm6
2.48	-2.95	Rnf181	4.32	-1.85	Tdrp
3.22	-2.92	Bhmt	2.22	-1.85	M1ap
3.36	-2.88	Ifitm3	4.38	-1.83	Mybl2
7.08	-2.75	Alpl	2.00	-1.83	Sept1
8.99	-2.74	Echdc2	4.79	-1.80	Atn1
8.06	-2.66	Cdyl2	4.44	-1.79	Bdh1
5.27	-2.65	Mov10	4.16	-1.79	Grsf1
4.85	-2.61	Gaa	6.04	-1.78	Snx18
2.72	-2.57	Ebp	2.79	-1.77	Nfu1
7.54	-2.51	Ing4	3.35	-1.74	Itga6
3.49	-2.51	Slc25a20	2.69	-1.73	Isoc2a
4.85	-2.51	Acot9	5.09	-1.72	Rpp25
3.83	-2.50	Acsf2	2.97	-1.71	Derl1
4.01	-2.50	Sod2	3.40	-1.70	Bap1
8.61	-2.45	Tdh	5.36	-1.70	Rfx2
3.66	-2.40	Zfp57	3.56	-1.70	Pdxk
5.19	-2.37	Znf710	4.08	-1.70	Gtf2h1
4.65	-2.36	Upp1	5.79	-1.70	Mpnd
6.32	-2.28	Gna13	2.80	-1.68	Cryz
6.29	-2.28	Tbx3	2.72	-1.66	Arap1
2.67	-2.27	Gabarapl2	7.73	-1.64	Stat3
9.55	-2.26	Sox2	2.71	-1.64	Mmgt1
12.43	-2.21	Nfatc2ip	3.26	-1.64	Tomm22
6.77	-2.17	Tfap2c	8.20	-1.63	Zmym3
2.70	-2.16	Lamtor1	4.64	-1.62	H2afy
5.90	-2.11	Chd2	4.01	-1.61	Nt5c3b
9.70	-2.11	Ckb	3.06	-1.61	Marc2
2.06	-2.10	Pam16	8.15	-1.61	Aff1
6.01	-2.09	Klf5	4.72	-1.59	Zswim1
1.36	-2.09	Ak4	10.94	-1.59	Kdm2b
4.63	-2.06	Ap1s1	2.80	-1.58	Cobl
3.56	-2.06	Hexb	5.77	-1.57	Pcyt1a
4.14	-2.05	Eps8l2	8.89	-1.57	Kat6b
7.30	-2.05	Arid3a	2.87	-1.57	Tmx3
3.52	-2.04	Ckmt1	2.23	-1.56	Cdipt
4.10	-2.02	Jup	7.10	-1.56	Lap3
2.20	-2.00	Tmem256	8.59	-1.56	Znf148



5.16	-1.54	Lmna	3.41	1.67	Kif1a
2.29	-1.54	Amt	4.93	1.71	Pfkm
2.60	-1.54	Zfp292	1.79	1.71	Hmgn2
6.96	-1.54	Stk38	4.46	1.72	Fads2
3.49	-1.53	Fbxo15	7.71	1.73	Mllt4
9.77	-1.52	Lactb2	3.32	1.73	Prmt3
2.74	-1.52	Adpgk	3.42	1.73	Cyp51a1
3.57	-1.52	Slc25a17	3.64	1.75	Nxn
2.46	-1.52	Sh3bp1	3.36	1.75	Crmp1
4.50	-1.51	Synj2bp	6.27	1.76	Nr2f6
4.36	-1.51	Anxa6	5.03	1.76	Znf532
2.64	-1.51	Ndufa9	2.86	1.81	Akap9
5.76	-1.48	Gpx1	3.73	1.81	Rlim
3.20	-1.47	Ulk1	1.85	1.82	Mapre2
3.86	-1.45	Smyd3	2.91	1.83	Col18a1
4.17	-1.45	Gtf2h4	2.57	1.83	Pdlim7
3.87	-1.45	Ass1	5.48	1.83	Pou3f1
3.96	-1.44	Ttc39b	3.37	1.86	Fam208b
8.64	-1.43	Uap1l1	6.71	1.87	Prdx6
4.84	-1.42	Dst	4.18	1.89	Esrp1
6.64	-1.40	Asns	3.90	1.94	Flnc
4.92	-1.40	Acadm	5.94	2.06	Dnmt3l
4.43	-1.40	Pigs	1.17	2.13	Sumo2
6.27	-1.35	Trim35	2.07	2.15	Uhrf1
8.51	-1.33	Kdm4b	5.30	2.21	Suv39h1
5.85	1.36	Eif4g2	1.14	2.23	Hist2h2ab
5.48	1.41	Acaca	5.76	2.26	Prph
5.11	1.44	Mvk	5.50	2.31	Ptbp2
3.79	1.46	Znf326	2.68	2.35	Pycr2
6.99	1.50	Utf1	2.52	2.42	Ubr2
7.39	1.53	Ubr7	3.83	2.45	Otx2;Otx1
2.61	1.53	Sumo3	11.10	2.62	Shroom2
3.25	1.53	Mafg	7.13	2.80	Myh10
2.41	1.53	Pnpo	6.60	2.81	Plcg2
3.40	1.53	Pdcd4	7.11	2.90	Vrtn
3.10	1.54	Arhgef2	2.22	2.91	Unc45a
5.45	1.56	Gclm	8.93	2.94	Grhl2
2.35	1.57	Adk	2.96	3.04	Hist3h2ba;Hist3h2bb
2.21	1.57	Cbx1	9.89	3.10	Ina
6.69	1.59	Chd9	3.16	3.27	Hmga1
2.68	1.60	Smc1b	7.63	3.46	Acsl4
6.86	1.60	Hmgcs1	5.49	3.66	Hist1h1t
2.24	1.61	Zic3	7.75	3.95	Cgn
2.73	1.62	Casp3	5.60	4.26	Sall2
4.73	1.65	Tjp1	9.36	4.55	Dnmt3b
2.57	1.66	Crip2	11.43	5.35	Dnmt3a

## 2. Results

**Supplementary Table 2: Proteins significantly correlating with "Dnmt-TKO low", "Tet-TKO low", or "Tdg-KO high" profiles in ESCs**

Perseus export of profile correlations; only proteins with a profile FDR < 0.02 (highlighted by grey background) for at least one of the profiles are shown.

Gene name	"Dnmt-TKO low" distance	"Tet-TKO low" distance	"Tdg-KO high" distance	"Dnmt-TKO low" FDR	"Tet-TKO low" FDR	"Tdg-KO high" FDR
1810009A15Rik	0.83	1.14	0.29	0.8482	1.0000	0.0003
2700029M09Rik	0.20	0.93	0.29	0.0074	0.8756	0.0003
3110082I17Rik	0.20	1.07	0.41	0.0074	1.0000	0.0041
A830080D01Rik	0.20	0.80	0.17	0.0074	0.4987	0.0000
Aatf	0.78	0.36	1.15	0.6366	0.0057	1.0000
Acin1	0.46	0.83	0.45	0.0327	0.5692	0.0094
Acly	1.20	1.22	0.47	1.0000	1.0000	0.0150
Actl6a	0.38	0.99	0.31	0.0138	1.0000	0.0003
Adnp2	0.55	0.99	0.27	0.0822	1.0000	0.0002
Adprh	0.32	1.48	1.28	0.0103	1.0000	1.0000
Aen	1.29	0.23	0.79	1.0000	0.0005	0.5954
Aff1	1.17	0.28	0.66	1.0000	0.0011	0.1973
Ahdc1	1.26	0.09	0.73	1.0000	0.0000	0.3849
Ak1	1.36	0.34	0.70	1.0000	0.0038	0.3033
Alyref	0.40	0.86	0.44	0.0156	0.6613	0.0075
Anln	0.93	0.39	0.42	1.0000	0.0096	0.0049
Anp32a	0.36	1.20	0.39	0.0125	1.0000	0.0021
Anp32e	0.39	0.87	0.32	0.0144	0.6944	0.0003
Apex1	0.65	1.09	0.39	0.2381	1.0000	0.0023
Api5	0.52	0.81	0.46	0.0630	0.5309	0.0107
Arid5b	1.33	0.36	0.61	1.0000	0.0051	0.1096
Armc6	1.61	0.40	1.33	1.0000	0.0105	1.0000
Asns	0.39	1.42	1.11	0.0152	1.0000	1.0000
Atad2	0.59	0.80	0.24	0.1295	0.5053	0.0000
Atad5	0.77	0.40	0.30	0.5834	0.0121	0.0003
Atf7ip	0.84	0.23	0.81	0.8628	0.0006	0.6637
Atm	1.05	0.64	0.28	1.0000	0.1756	0.0003
Atox1	0.82	0.70	0.24	0.8112	0.2816	0.0000
Atp2a2	0.70	1.46	0.44	0.3644	1.0000	0.0072
Atr	1.14	0.33	0.46	1.0000	0.0029	0.0108
Aurka	1.56	0.32	1.19	1.0000	0.0021	1.0000
Aurkb	0.84	0.21	0.55	0.8644	0.0000	0.0512
BC055324	0.76	0.87	0.42	0.5693	0.6980	0.0050
Bop1	1.32	0.31	1.35	1.0000	0.0020	1.0000
Brca1	0.52	0.76	0.37	0.0622	0.3933	0.0013
Brd2	0.39	0.72	0.46	0.0144	0.3124	0.0114
Brd3	0.41	0.70	0.28	0.0162	0.2902	0.0003
Bud13	0.95	0.35	1.22	1.0000	0.0038	1.0000
Bzw2	1.40	0.23	0.74	1.0000	0.0006	0.3920
Cactin	1.11	0.22	1.03	1.0000	0.0005	1.0000
Cand1	0.82	0.80	0.18	0.8053	0.5045	0.0000
Cbfa2t2	0.82	0.55	0.39	0.8218	0.0739	0.0023
Cbx3	0.40	1.05	0.39	0.0160	1.0000	0.0023
Cbx5	0.25	1.16	0.54	0.0077	1.0000	0.0425
Ccar1	0.33	0.73	0.48	0.0102	0.3334	0.0162
Ccar2	1.40	0.10	0.93	1.0000	0.0000	1.0000

Ccnk	0.47	0.74	0.28	0.0384	0.3599	0.0003
Cdc27	1.18	0.30	0.73	1.0000	0.0015	0.3763
Cdc42	1.01	1.32	0.44	1.0000	1.0000	0.0086
Cdc5l	0.53	0.70	0.30	0.0714	0.2812	0.0004
Cdca8	1.09	0.27	0.51	1.0000	0.0010	0.0270
Cdk12	0.79	0.42	0.26	0.6848	0.0142	0.0002
Cdyl2	0.11	1.10	0.90	0.0117	1.0000	1.0000
Cebpz	0.24	0.84	0.93	0.0082	0.5958	1.0000
Cenpv	0.30	1.19	0.66	0.0095	1.0000	0.2028
Cfdp1	0.33	0.94	0.47	0.0100	0.8882	0.0132
Chaf1a	0.37	1.01	0.40	0.0133	1.0000	0.0023
Champ1	0.25	0.99	0.39	0.0076	0.9990	0.0020
Chd1	0.86	0.96	0.25	0.9123	0.9320	0.0000
Chd2	0.67	0.84	0.26	0.2968	0.5993	0.0001
Chd4	0.53	0.58	0.48	0.0699	0.0981	0.0153
Chd8	1.65	0.38	1.01	1.0000	0.0084	1.0000
Chek1	1.12	0.99	0.42	1.0000	1.0000	0.0054
Ckap2	1.06	0.27	1.05	1.0000	0.0009	1.0000
Ckap5	1.16	0.50	0.48	1.0000	0.0448	0.0156
Clic4	0.42	1.37	0.68	0.0198	1.0000	0.2469
Clk4	0.71	0.41	0.52	0.4092	0.0126	0.0341
Cndp2	0.38	1.74	0.97	0.0133	1.0000	1.0000
Cnn3	0.29	1.48	0.85	0.0085	1.0000	0.8027
Coasy	0.74	0.52	0.33	0.4924	0.0507	0.0004
Coil	0.98	0.23	0.68	1.0000	0.0005	0.2388
Cpsf2	0.41	0.57	0.76	0.0177	0.0883	0.4917
Crip2	0.94	0.75	0.19	1.0000	0.3734	0.0000
Cse1l	0.31	0.73	0.44	0.0098	0.3340	0.0074
Csnk2a2	1.34	0.39	1.49	1.0000	0.0094	1.0000
Ctc1	0.85	0.29	1.00	0.8946	0.0011	1.0000
Cul4b	0.84	0.71	0.26	0.8475	0.3022	0.0000
Cwc22	0.30	0.97	0.53	0.0091	0.9529	0.0380
Cxxc1	1.37	0.25	1.16	1.0000	0.0008	1.0000
Dazap1	0.23	1.41	0.93	0.0074	1.0000	1.0000
Dcun1d5	1.57	0.37	0.94	1.0000	0.0073	1.0000
Ddb1	0.79	1.44	0.45	0.6599	1.0000	0.0088
Ddx1	0.67	0.83	0.29	0.2809	0.5918	0.0003
Ddx10	1.10	0.30	1.33	1.0000	0.0014	1.0000
Ddx21	0.62	0.34	0.71	0.1930	0.0033	0.3290
Ddx46	0.54	0.73	0.30	0.0787	0.3268	0.0004
Ddx47	0.36	0.72	0.60	0.0126	0.3254	0.1047
Ddx5	0.42	0.67	1.08	0.0177	0.2247	1.0000
Ddx50	0.62	0.29	0.76	0.1727	0.0013	0.4848
Dek	0.45	0.97	0.34	0.0308	0.9677	0.0006
Dgcr8	1.56	0.24	1.05	1.0000	0.0006	1.0000
Dhx36	1.34	0.18	0.83	1.0000	0.0000	0.7166
Dhx8	1.11	0.26	0.84	1.0000	0.0009	0.7725
Dis3	1.09	0.51	0.42	1.0000	0.0499	0.0050
Dmap1	1.06	0.22	0.57	1.0000	0.0000	0.0683
Dnajc8	0.32	0.96	0.45	0.0099	0.9531	0.0099
Dnajc9	0.74	0.68	0.22	0.5106	0.2470	0.0000
Dnmt1	0.14	1.00	0.53	0.0100	1.0000	0.0343
Dnmt3a	0.20	1.21	0.73	0.0079	1.0000	0.3886
Dnmt3b	0.13	1.20	0.87	0.0100	1.0000	0.9044

## 2. Results

Dpf2	1.11	0.25	0.55	1.0000	0.0007	0.0515
Ect2	1.14	0.32	0.49	1.0000	0.0024	0.0179
Eed	0.48	0.90	0.33	0.0391	0.7603	0.0004
Eif1ax	0.64	1.02	0.16	0.2070	1.0000	0.0000
Eif4a3	0.22	1.22	0.91	0.0072	1.0000	1.0000
Eml2	0.37	1.30	1.20	0.0132	1.0000	1.0000
Engase	0.87	0.18	0.63	0.9637	0.0000	0.1347
Ercc4	0.20	1.19	0.94	0.0069	1.0000	1.0000
Esco2	0.75	0.75	0.29	0.5351	0.3662	0.0003
Esf1	0.54	0.74	0.33	0.0786	0.3460	0.0005
Esrrb	0.54	0.67	0.24	0.0772	0.2370	0.0000
Fabp3	0.36	1.55	0.72	0.0126	1.0000	0.3624
Fabp5	1.31	0.44	1.45	1.0000	0.0197	1.0000
Fam111a	0.70	0.89	0.27	0.3649	0.7459	0.0003
Fam169a	0.68	0.60	0.22	0.3205	0.1220	0.0000
Fam192a	0.97	0.31	0.87	1.0000	0.0016	0.8803
Fam208a	0.21	0.87	0.59	0.0069	0.6969	0.0826
Fam50a	0.30	0.83	0.48	0.0091	0.5808	0.0171
Fam98b	0.80	0.81	0.26	0.7398	0.5372	0.0003
Fanca	0.41	0.49	0.75	0.0170	0.0368	0.4496
Fancd2	0.55	0.74	0.41	0.0813	0.3658	0.0041
Fanci	0.53	0.59	0.24	0.0732	0.1041	0.0000
Fbl	0.30	1.12	1.20	0.0093	1.0000	1.0000
Figl1	1.05	0.28	0.88	1.0000	0.0009	0.9290
Fip1l1	0.26	0.89	0.99	0.0078	0.7336	1.0000
Fiz1	0.39	0.84	0.44	0.0151	0.6130	0.0078
Fkbp5	0.72	1.23	0.41	0.4281	1.0000	0.0037
Foxk2	1.18	0.23	0.87	1.0000	0.0006	0.8575
Ftsj3	0.50	0.44	1.03	0.0472	0.0198	1.0000
Fus	0.19	1.09	0.55	0.0081	1.0000	0.0522
Gapdh	0.55	1.27	0.39	0.0816	1.0000	0.0021
Gatad1	0.76	0.21	0.53	0.5584	0.0000	0.0341
Gatad2b	0.99	0.30	0.70	1.0000	0.0015	0.2950
Gle1	1.16	0.40	1.45	1.0000	0.0121	1.0000
Glr3	0.51	0.98	0.35	0.0558	0.9667	0.0008
Glyr1	0.26	0.96	0.52	0.0076	0.9355	0.0333
Gmps	1.05	0.89	0.21	1.0000	0.7596	0.0000
Gnl3	0.48	1.02	0.45	0.0388	1.0000	0.0101
Gnl3l	0.62	0.55	0.43	0.1768	0.0759	0.0057
Grwd1	1.66	0.44	1.36	1.0000	0.0199	1.0000
Gtf2e2	0.45	1.09	0.40	0.0289	1.0000	0.0034
Gtf2f2	0.36	0.95	0.41	0.0125	0.9187	0.0037
Gtf3c1	1.05	0.27	0.96	1.0000	0.0010	1.0000
Gtf3c2	0.94	0.27	1.07	1.0000	0.0009	1.0000
H2afy	0.25	1.02	0.45	0.0075	1.0000	0.0088
H2afy2	0.12	1.17	0.88	0.0100	1.0000	0.9266
Hat1	1.35	0.36	1.22	1.0000	0.0056	1.0000
Hdac1	0.30	1.27	0.68	0.0092	1.0000	0.2422
Hdac2	0.35	0.79	0.48	0.0127	0.4738	0.0162
Hdgf	0.41	1.01	0.27	0.0163	1.0000	0.0002
Hira	0.46	0.57	0.48	0.0342	0.0916	0.0172
Hist1h1a	0.23	1.09	0.45	0.0085	1.0000	0.0098
Hist1h1b	0.31	0.95	0.39	0.0093	0.9227	0.0020
Hist1h1c	0.26	1.08	0.42	0.0075	1.0000	0.0053

Hist1h1d	0.28	1.07	0.44	0.0086	1.0000	0.0081
Hist1h1e	0.44	0.95	0.30	0.0236	0.9141	0.0003
Hist1h2ab/h/f/k	1.37	0.27	1.32	1.0000	0.0009	1.0000
Hist1h2bb	0.41	0.91	0.48	0.0167	0.7967	0.0168
Hist2h2aa1;Hist2h2ac	0.86	0.40	1.12	0.9582	0.0107	1.0000
Hist2h2ab	0.88	0.81	0.46	1.0000	0.5379	0.0124
Hmces	1.23	0.43	0.43	1.0000	0.0187	0.0060
Hmgb1	0.45	1.09	0.37	0.0291	1.0000	0.0012
Hmgn2	0.94	0.79	0.10	1.0000	0.4891	0.0000
Hnrnpa0	0.20	1.27	0.83	0.0073	1.0000	0.7246
Hnrnpa3	0.18	1.40	0.93	0.0100	1.0000	1.0000
Hnrnpab	0.21	1.13	0.44	0.0068	1.0000	0.0076
Hnrnpd	0.57	1.03	0.31	0.1069	1.0000	0.0003
Hnrnpb2	1.49	0.33	1.15	1.0000	0.0025	1.0000
Hnrnp1	0.08	1.17	0.79	0.0100	1.0000	0.5867
Hnrnpm	0.26	1.06	1.17	0.0074	1.0000	1.0000
Hp1bp3	0.57	0.60	0.27	0.1075	0.1153	0.0002
Hspa4	0.59	1.27	0.22	0.1392	1.0000	0.0000
Huwe1	1.71	0.42	0.95	1.0000	0.0143	1.0000
Ice1	0.99	0.33	0.99	1.0000	0.0030	1.0000
Ik	0.24	0.90	0.46	0.0082	0.7792	0.0111
Impact	1.34	0.15	1.05	1.0000	0.0000	1.0000
Incenp	0.87	0.51	0.36	0.9715	0.0464	0.0012
Ing4	0.91	0.24	1.03	1.0000	0.0005	1.0000
Ints1	1.06	0.19	0.96	1.0000	0.0000	1.0000
Ints2	1.04	0.41	1.46	1.0000	0.0135	1.0000
Ints5	1.23	0.41	1.51	1.0000	0.0135	1.0000
Ints9	1.01	0.22	0.79	1.0000	0.0005	0.5773
Ipo8	1.60	0.43	0.85	1.0000	0.0184	0.7865
Irf2bp2	0.97	0.29	0.46	1.0000	0.0012	0.0112
Iws1	0.39	1.01	0.28	0.0145	1.0000	0.0002
Kansl1	0.94	0.23	0.85	1.0000	0.0005	0.7859
Kansl3	1.07	0.24	0.69	1.0000	0.0006	0.2556
Kat6b	0.48	0.65	0.47	0.0381	0.1955	0.0142
Kat7	0.53	0.64	0.47	0.0719	0.1702	0.0141
Kdm1b	0.34	1.00	0.54	0.0112	1.0000	0.0459
Kiaa1429	0.67	0.49	0.33	0.2759	0.0379	0.0005
Kif15	1.26	0.29	1.27	1.0000	0.0014	1.0000
Kif18b	0.94	0.17	0.67	1.0000	0.0000	0.2210
Kif20b	1.05	0.32	0.38	1.0000	0.0024	0.0018
Kif22	0.77	0.30	0.34	0.5941	0.0014	0.0005
Kifc1	1.28	0.41	1.51	1.0000	0.0122	1.0000
Klf4	0.40	0.52	0.91	0.0158	0.0536	1.0000
Kmt2a	1.21	0.17	0.73	1.0000	0.0000	0.3646
Lig1	0.96	0.77	0.19	1.0000	0.4234	0.0000
Lig3	0.40	0.98	0.42	0.0154	0.9687	0.0053
Lmnbl	0.09	1.18	0.93	0.0125	1.0000	1.0000
Lmnbl2	0.32	0.62	0.85	0.0102	0.1417	0.7843
Lrrc40	1.39	0.22	1.03	1.0000	0.0000	1.0000
LRWD1	0.59	0.50	0.31	0.1366	0.0388	0.0003
Lta4h	1.25	0.72	0.31	1.0000	0.3164	0.0003
Mcm3ap	0.66	0.39	0.47	0.2752	0.0091	0.0130
Mcm4	0.23	0.99	0.38	0.0073	1.0000	0.0020
Mcm5	0.31	0.85	0.48	0.0095	0.6261	0.0162

## 2. Results

Mdc1	0.78	0.29	0.81	0.6284	0.0014	0.6328
Mecp2	0.58	0.61	0.28	0.1152	0.1405	0.0002
Med1	1.15	0.62	0.37	1.0000	0.1493	0.0014
Med24	1.41	0.16	0.90	1.0000	0.0000	1.0000
Mettl16	1.24	0.09	0.69	1.0000	0.0000	0.2779
Mfap1	0.22	0.92	0.58	0.0071	0.8338	0.0795
Mga	0.72	0.81	0.31	0.4176	0.5314	0.0003
Mina	1.66	0.19	0.95	1.0000	0.0000	1.0000
Mms22l	1.41	0.72	0.41	1.0000	0.3147	0.0037
Mpg	0.93	0.34	0.83	1.0000	0.0032	0.7104
Mre11a	0.78	0.37	0.68	0.6280	0.0068	0.2399
Msh2	0.36	0.98	0.23	0.0124	0.9681	0.0000
Msh6	0.45	0.99	0.33	0.0290	1.0000	0.0005
Mta1	1.02	0.24	0.41	1.0000	0.0005	0.0037
Mtf2	0.27	1.28	0.62	0.0084	1.0000	0.1347
Mybbp1a	0.39	0.49	0.77	0.0145	0.0346	0.5113
Mybl2	0.53	1.04	0.34	0.0695	1.0000	0.0006
Myef2	1.26	0.42	1.26	1.0000	0.0142	1.0000
Ncapg2	0.53	0.76	0.36	0.0720	0.3937	0.0010
Ncbp3	0.37	0.69	1.61	0.0130	0.2683	1.0000
Ncl	0.34	1.15	0.32	0.0119	1.0000	0.0003
Ncor2	1.39	0.12	1.00	1.0000	0.0000	1.0000
Nde1	1.34	0.22	1.17	1.0000	0.0000	1.0000
Nelfa	0.56	0.86	0.24	0.1066	0.6588	0.0000
Nelfcd	0.52	0.86	0.30	0.0653	0.6577	0.0004
Nfatc2ip	0.53	0.93	0.30	0.0708	0.8726	0.0004
Nle1	1.35	0.26	0.69	1.0000	0.0008	0.2645
Nmral1	1.59	0.25	1.09	1.0000	0.0008	1.0000
Noc2l	0.35	0.77	1.07	0.0117	0.4185	1.0000
Nol10	0.34	0.65	0.69	0.0104	0.1868	0.2657
Nol11	0.95	0.18	0.79	1.0000	0.0000	0.5873
Nol8	1.01	0.37	1.41	1.0000	0.0068	1.0000
Nom1	1.27	0.33	1.32	1.0000	0.0027	1.0000
Nono	0.21	1.22	1.02	0.0070	1.0000	1.0000
Nop9	1.25	0.23	0.85	1.0000	0.0004	0.8144
Npm1	0.24	1.18	0.48	0.0079	1.0000	0.0167
Nrde2	1.08	0.23	0.80	1.0000	0.0005	0.5950
Nsun2	1.46	0.48	0.39	1.0000	0.0310	0.0023
Ntmt1	1.55	0.41	1.12	1.0000	0.0141	1.0000
Nup133	0.89	0.27	1.00	1.0000	0.0009	1.0000
Nup188	0.89	0.33	1.20	1.0000	0.0029	1.0000
Nup43	0.92	0.37	1.09	1.0000	0.0068	1.0000
Nup50	0.36	0.73	0.36	0.0125	0.3332	0.0011
Nup93	0.39	0.74	1.30	0.0152	0.3628	1.0000
Nup98	0.76	0.30	1.00	0.5747	0.0014	1.0000
Ogt	0.94	0.11	0.80	1.0000	0.0000	0.6227
Orc1	0.70	0.33	0.46	0.3652	0.0027	0.0124
Orc2	0.26	0.87	0.36	0.0073	0.6972	0.0010
Orc3	0.41	0.59	0.31	0.0164	0.1112	0.0003
Orc4	0.53	0.54	0.37	0.0702	0.0678	0.0014
Orc5	0.48	0.62	0.31	0.0382	0.1412	0.0003
Papd4	1.17	0.21	0.90	1.0000	0.0000	0.9953
Papola	1.65	0.38	0.65	1.0000	0.0086	0.1754
Papolg	0.84	0.62	0.41	0.8781	0.1414	0.0040

Parn	0.40	0.96	0.49	0.0156	0.9533	0.0205
Parp1	0.49	1.08	0.31	0.0469	1.0000	0.0004
Paxbp1	0.77	0.48	0.42	0.6202	0.0298	0.0042
Pbrm1	0.33	0.99	0.34	0.0099	1.0000	0.0005
Pcid2	0.38	0.59	0.56	0.0145	0.1042	0.0608
Pcna	0.67	1.19	0.22	0.2818	1.0000	0.0000
Pcnp	0.65	1.14	0.40	0.2432	1.0000	0.0030
Pcyt1a	1.63	0.35	1.32	1.0000	0.0049	1.0000
Pdcd11	0.63	0.32	0.95	0.2008	0.0021	1.0000
Pds5a	0.37	0.79	0.45	0.0131	0.4921	0.0104
Pds5b	0.43	0.82	0.23	0.0203	0.5384	0.0000
Pebp1	0.95	0.31	0.53	1.0000	0.0016	0.0345
Pfas	1.53	0.62	0.38	1.0000	0.1473	0.0018
Pgk1	0.71	0.97	0.28	0.3965	0.9621	0.0002
Phc1	1.33	0.25	0.89	1.0000	0.0007	0.9915
Phf20	1.04	0.20	0.90	1.0000	0.0000	0.9943
Phf23	0.63	0.39	0.51	0.2003	0.0089	0.0269
Phip	0.68	0.57	0.27	0.3135	0.0873	0.0002
Pias2	0.61	0.74	0.34	0.1676	0.3459	0.0007
Pkn2	1.71	0.38	0.77	1.0000	0.0088	0.5422
Plk1	0.88	0.41	0.45	1.0000	0.0132	0.0095
Plrg1	0.89	0.29	0.89	1.0000	0.0011	0.9490
Pogz	0.36	0.76	0.59	0.0122	0.3941	0.0819
Pola1	0.72	0.96	0.34	0.4224	0.9529	0.0006
Poldip3	1.10	0.27	0.83	1.0000	0.0009	0.7142
Polr1e	1.07	0.14	0.86	1.0000	0.0000	0.8502
Polr2b	1.25	0.50	0.34	1.0000	0.0384	0.0006
Pot1	0.80	0.44	0.68	0.7084	0.0199	0.2472
Pou5f1	0.40	0.73	0.64	0.0158	0.3332	0.1759
Ppid	0.58	0.90	0.41	0.1143	0.7874	0.0036
Ppig	0.40	0.68	0.31	0.0156	0.2512	0.0003
Ppm1g	0.48	0.88	0.30	0.0413	0.7169	0.0004
Ppp2r1b	0.93	0.42	0.27	1.0000	0.0158	0.0002
Ppp2r5c	1.11	0.36	0.68	1.0000	0.0066	0.2424
Ppwd1	0.85	0.30	0.50	0.8797	0.0016	0.0209
Prkdc	1.14	0.23	0.43	1.0000	0.0005	0.0072
Prpf31	0.67	0.35	0.61	0.2972	0.0038	0.1091
Prpf38b	0.52	0.69	0.45	0.0628	0.2528	0.0088
Prpf40a	0.57	0.58	0.45	0.1071	0.0946	0.0088
Prpf4b	0.73	0.45	0.29	0.4518	0.0212	0.0003
Prpf6	0.30	0.78	0.60	0.0092	0.4508	0.1008
Prpf8	0.68	0.34	0.49	0.3108	0.0030	0.0206
Psip1	0.33	1.04	0.37	0.0103	1.0000	0.0017
Psmc3ip	0.42	0.57	0.62	0.0194	0.0893	0.1331
Psmc5	1.65	0.41	0.64	1.0000	0.0135	0.1703
Pspc1	1.22	0.12	0.74	1.0000	0.0000	0.4135
Psrc1	1.60	0.32	1.40	1.0000	0.0021	1.0000
Ptbp1	0.31	0.89	0.45	0.0095	0.7495	0.0088
Ptbp2	1.15	0.63	0.27	1.0000	0.1666	0.0002
Ptma	0.85	1.33	0.23	0.8941	1.0000	0.0000
Puf60	0.62	0.83	0.33	0.1703	0.5740	0.0005
Pus1	0.78	0.83	0.37	0.6349	0.5742	0.0013
Qrich1	0.76	0.66	0.23	0.5525	0.1987	0.0000
Racgap1	0.61	0.94	0.35	0.1566	0.8902	0.0009



## 2. Results

Rbbp7	0.37	1.14	0.54	0.0131	1.0000	0.0418
Rbbp8	0.98	0.27	0.74	1.0000	0.0009	0.3979
Rbm12b2	0.41	0.69	0.52	0.0169	0.2549	0.0311
Rbm19	1.10	0.32	1.13	1.0000	0.0022	1.0000
Rbm27	0.97	0.36	0.51	1.0000	0.0056	0.0261
Rbm39	0.65	0.38	0.65	0.2494	0.0084	0.1750
Rbmxl1	0.34	1.23	1.12	0.0103	1.0000	1.0000
Rcc1	0.54	0.65	0.26	0.0765	0.1869	0.0003
Rcl1	1.07	0.42	1.44	1.0000	0.0174	1.0000
Rfc2	0.18	1.05	0.54	0.0095	1.0000	0.0452
Rfc5	0.88	0.26	0.62	1.0000	0.0008	0.1271
Rif1	0.22	0.86	0.45	0.0068	0.6607	0.0094
Rnf169	0.59	0.44	0.41	0.1384	0.0187	0.0041
Rnh1	0.52	1.27	0.35	0.0653	1.0000	0.0006
Rnps1	0.59	0.77	0.26	0.1360	0.4366	0.0002
Rpa1	0.32	1.09	0.35	0.0103	1.0000	0.0006
Rpa2	0.28	1.23	0.69	0.0084	1.0000	0.2622
Rprd2	1.32	0.32	0.75	1.0000	0.0020	0.4445
Rps6ka3	1.37	0.74	0.32	1.0000	0.3507	0.0003
Rpusd2	1.68	0.40	0.92	1.0000	0.0111	1.0000
Rsl1d1	0.16	1.17	0.70	0.0106	1.0000	0.2806
Rtcb	0.78	0.87	0.23	0.6303	0.7006	0.0000
Ruvbl1	0.51	0.83	0.47	0.0556	0.5696	0.0142
Rxrb	1.64	0.38	1.10	1.0000	0.0080	1.0000
Sae1	0.49	1.31	0.37	0.0472	1.0000	0.0013
Sall4	0.57	0.64	0.30	0.1071	0.1744	0.0002
Sap130	0.99	0.25	1.10	1.0000	0.0008	1.0000
Sap30bp	0.48	0.99	0.35	0.0398	1.0000	0.0008
Scaf1	1.27	0.39	1.30	1.0000	0.0104	1.0000
Sde2	0.87	0.50	0.24	0.9600	0.0407	0.0000
Senp1	0.99	0.35	1.15	1.0000	0.0048	1.0000
Senp3	0.25	0.78	0.96	0.0080	0.4393	1.0000
Set	0.57	1.18	0.31	0.1132	1.0000	0.0003
Setd6	0.94	0.31	1.10	1.0000	0.0016	1.0000
Sf3b1	0.40	0.58	0.71	0.0159	0.0969	0.3115
Sgol2	0.84	0.58	0.31	0.8503	0.0955	0.0003
Shprh	1.03	0.35	0.55	1.0000	0.0039	0.0476
Sin3a	0.42	0.53	0.42	0.0198	0.0602	0.0043
Sirt6	0.90	0.37	0.35	1.0000	0.0068	0.0008
Skiv2l2	0.55	0.95	0.36	0.0810	0.9053	0.0010
Smarca5	0.15	1.06	0.58	0.0093	1.0000	0.0809
Smarcad1	0.49	0.76	0.28	0.0422	0.3896	0.0003
Smarcd1	1.33	0.28	0.82	1.0000	0.0009	0.7034
Smc1a	0.32	1.08	0.44	0.0101	1.0000	0.0072
Smc2	0.80	0.63	0.21	0.6880	0.1552	0.0000
Smc3	0.48	0.59	0.35	0.0381	0.1051	0.0008
Smc5	0.38	1.22	0.47	0.0131	1.0000	0.0131
Smek1	1.59	0.42	0.71	1.0000	0.0143	0.3156
Smpd4	0.83	0.33	1.05	0.8342	0.0027	1.0000
Snrnp200	0.53	0.39	0.46	0.0721	0.0096	0.0106
Sox2	1.29	0.32	0.58	1.0000	0.0022	0.0764
Spag7	0.73	0.74	0.34	0.4821	0.3462	0.0005
Srbd1	0.78	0.28	0.59	0.6200	0.0010	0.0826
Srrm1	0.39	0.94	0.25	0.0153	0.8902	0.0000

Ssb	0.40	1.01	0.30	0.0158	1.0000	0.0003
Ssrp1	0.37	0.72	0.33	0.0131	0.3241	0.0005
Stag1	0.32	0.93	0.53	0.0099	0.8750	0.0395
Stag2	1.02	0.39	0.34	1.0000	0.0106	0.0006
Stag3	0.78	0.44	0.48	0.6226	0.0198	0.0153
Stat3	0.76	0.27	0.53	0.5776	0.0009	0.0350
Stip1	0.52	1.42	0.43	0.0619	1.0000	0.0055
Stx8	1.64	0.34	1.04	1.0000	0.0032	1.0000
Sumo2	0.63	1.28	0.43	0.2057	1.0000	0.0055
Sumo3	0.96	0.89	0.17	1.0000	0.7328	0.0000
Supt16h	0.32	0.90	0.29	0.0101	0.7874	0.0003
Supt5h	0.81	1.04	0.46	0.7506	1.0000	0.0112
Suv420h2	0.73	0.25	0.72	0.4694	0.0007	0.3512
Suz12	0.24	1.00	0.50	0.0085	1.0000	0.0220
Sympk	0.40	0.59	0.57	0.0157	0.1084	0.0693
Taf1	1.02	0.29	0.49	1.0000	0.0013	0.0178
Taf6	0.95	0.23	0.50	1.0000	0.0005	0.0249
Tcea3	1.32	0.19	0.63	1.0000	0.0000	0.1520
Tcf20	1.35	0.20	0.82	1.0000	0.0000	0.7024
Tdp1	0.76	1.19	0.35	0.5781	1.0000	0.0008
Tdp2	0.72	0.69	0.35	0.4181	0.2682	0.0008
Terf2	0.25	0.82	0.48	0.0076	0.5480	0.0179
Terf2ip	0.65	0.73	0.19	0.2276	0.3409	0.0000
Tet1	1.05	0.15	0.62	1.0000	0.0000	0.1319
Tet2	0.60	0.43	0.83	0.1448	0.0183	0.7146
Tfcp2l1	0.28	1.31	0.73	0.0084	1.0000	0.3653
Thoc1	0.37	0.90	0.26	0.0131	0.7862	0.0001
Thoc2	0.38	0.82	0.27	0.0134	0.5589	0.0003
Thyn1	0.36	1.17	0.44	0.0122	1.0000	0.0081
Tkt	0.85	1.56	0.39	0.9090	1.0000	0.0023
Tle3	1.01	0.23	0.83	1.0000	0.0005	0.7095
Tmem209	0.61	0.40	1.02	0.1588	0.0121	1.0000
Tmpo	0.29	0.97	0.74	0.0092	0.9643	0.4046
Tomm34	1.36	0.38	0.47	1.0000	0.0078	0.0146
Top1	0.41	1.16	0.20	0.0166	1.0000	0.0000
Top2a	0.58	0.65	0.28	0.1193	0.1903	0.0002
Top2b	0.92	0.76	0.06	1.0000	0.3932	0.0000
Tp53	0.94	0.31	0.89	1.0000	0.0016	0.9746
Tpr	0.45	0.73	0.36	0.0280	0.3422	0.0011
Tpx2	0.62	0.46	0.31	0.1742	0.0266	0.0003
Tra2b	0.37	0.82	1.08	0.0132	0.5564	1.0000
Traip	1.11	0.11	0.72	1.0000	0.0000	0.3439
Trappc11	1.17	0.32	0.88	1.0000	0.0020	0.9462
Trim24	0.19	1.05	0.40	0.0086	1.0000	0.0030
Trim28	0.24	0.86	0.43	0.0079	0.6565	0.0064
Trim33	0.42	0.94	0.40	0.0197	0.9007	0.0030
Trrap	1.00	0.24	0.83	1.0000	0.0006	0.7431
Tsr1	0.83	0.25	0.57	0.8372	0.0007	0.0670
Ttc27	1.59	0.41	1.26	1.0000	0.0126	1.0000
Ttk	1.58	0.34	0.64	1.0000	0.0038	0.1663
Tuba4a	1.14	0.43	1.31	1.0000	0.0184	1.0000
U2af1	0.70	0.40	0.79	0.3813	0.0106	0.5744
U2af2	0.45	0.66	0.34	0.0287	0.2056	0.0006
Uba2	0.85	1.18	0.31	0.9023	1.0000	0.0003

## 2. Results

Ubr5	1.57	0.29	1.07	1.0000	0.0011	1.0000
Ubr7	1.00	0.98	0.31	1.0000	0.9717	0.0003
Ubtf	0.30	1.10	0.33	0.0093	1.0000	0.0005
Ubxn7	0.62	0.82	0.34	0.1737	0.5439	0.0006
Upp1	1.23	0.22	0.68	1.0000	0.0005	0.2423
Usp28	0.84	0.41	0.38	0.8677	0.0135	0.0019
Usp39	0.64	0.43	0.45	0.2157	0.0183	0.0088
Usp5	1.03	0.65	0.46	1.0000	0.1901	0.0117
Usp7	0.65	0.81	0.32	0.2426	0.5305	0.0003
Utp14a	0.96	0.13	0.59	1.0000	0.0000	0.0878
Utp20	0.93	0.10	0.79	1.0000	0.0000	0.5704
Vim	1.50	0.43	1.03	1.0000	0.0177	1.0000
Vrk1	0.40	0.93	0.30	0.0153	0.8667	0.0004
Vwa9	0.46	0.79	0.39	0.0310	0.4774	0.0023
Wapal	0.69	0.94	0.47	0.3358	0.9016	0.0149
Wdhd1	0.81	0.48	0.46	0.7699	0.0305	0.0104
Wdr33	1.32	0.40	1.48	1.0000	0.0110	1.0000
Wdr5	0.29	0.98	0.36	0.0088	0.9687	0.0012
Wdr55	1.52	0.25	1.07	1.0000	0.0008	1.0000
Xpo1	0.40	0.72	0.53	0.0155	0.3191	0.0379
Xpo5	0.38	0.52	0.67	0.0141	0.0545	0.2321
Xpot	1.23	0.27	0.96	1.0000	0.0009	1.0000
Xrcc1	0.27	0.86	0.38	0.0084	0.6544	0.0018
Xrcc5	0.45	1.05	0.39	0.0280	1.0000	0.0022
Xrcc6	0.86	0.57	0.19	0.9549	0.0926	0.0000
Ywhaq	0.78	1.29	0.32	0.6545	1.0000	0.0004
Ywhaz	0.46	1.40	0.44	0.0310	1.0000	0.0072
Zfp292	0.56	0.91	0.26	0.0943	0.8070	0.0003
Zfp809	1.24	0.39	0.58	1.0000	0.0092	0.0745
Zmiz2	1.29	0.25	1.04	1.0000	0.0008	1.0000
Zmym2	0.98	0.37	0.27	1.0000	0.0066	0.0002
Znf106	0.51	0.37	0.50	0.0560	0.0068	0.0220
Znf280c	1.12	0.34	0.33	1.0000	0.0032	0.0004
Znf512	0.25	0.95	0.33	0.0078	0.9038	0.0003
Znf516	0.32	0.90	1.09	0.0102	0.7769	1.0000

### 3. Discussion

#### 3.1 Regulation of TET proteins by interaction partners and PTMs

##### 3.1.1 TET proteins, base excision repair, and DNA demethylation

The cellular function of proteins not only depends on their catalytic activity and structure, but is often regulated by protein-protein-interactions, the availability of co-factors, PTMs, and intracellular localization. This thesis aims to characterize the regulation of TET proteins by identification and analysis of interacting proteins as well as phosphorylation and *O*-GlcNAcylation sites.

In a first screen for protein-protein interactions of TET1, three members of the BER pathway co-precipitated with TET1: PARP1, LIG3, and XRCC1 (Muller *et al.*, 2014). This prompted us to evaluate other members of the BER machinery for their interaction potential with TET proteins. BER is a multi-step process, which consists of the following parts: excision of the damaged or mispaired base by a glycosylase, incision, end processing, gap filling and ligation (Krokan and Bjoras, 2013). In addition to LIG3, PARP1, and XRCC1, we identified five DNA glycosylases, namely TDG, MBD4, NEIL1, NEIL2, and NEIL3, as interaction partners of all three TETs (Figure 7). With the exception of XRCC1, all detected interactions are completely independent of TET catalytic activity. For XRCC1, we observe a cooperative effect between direct protein-protein interaction and possible recruitment by the generated oxidized cytosine variants (Muller *et al.*, 2014).

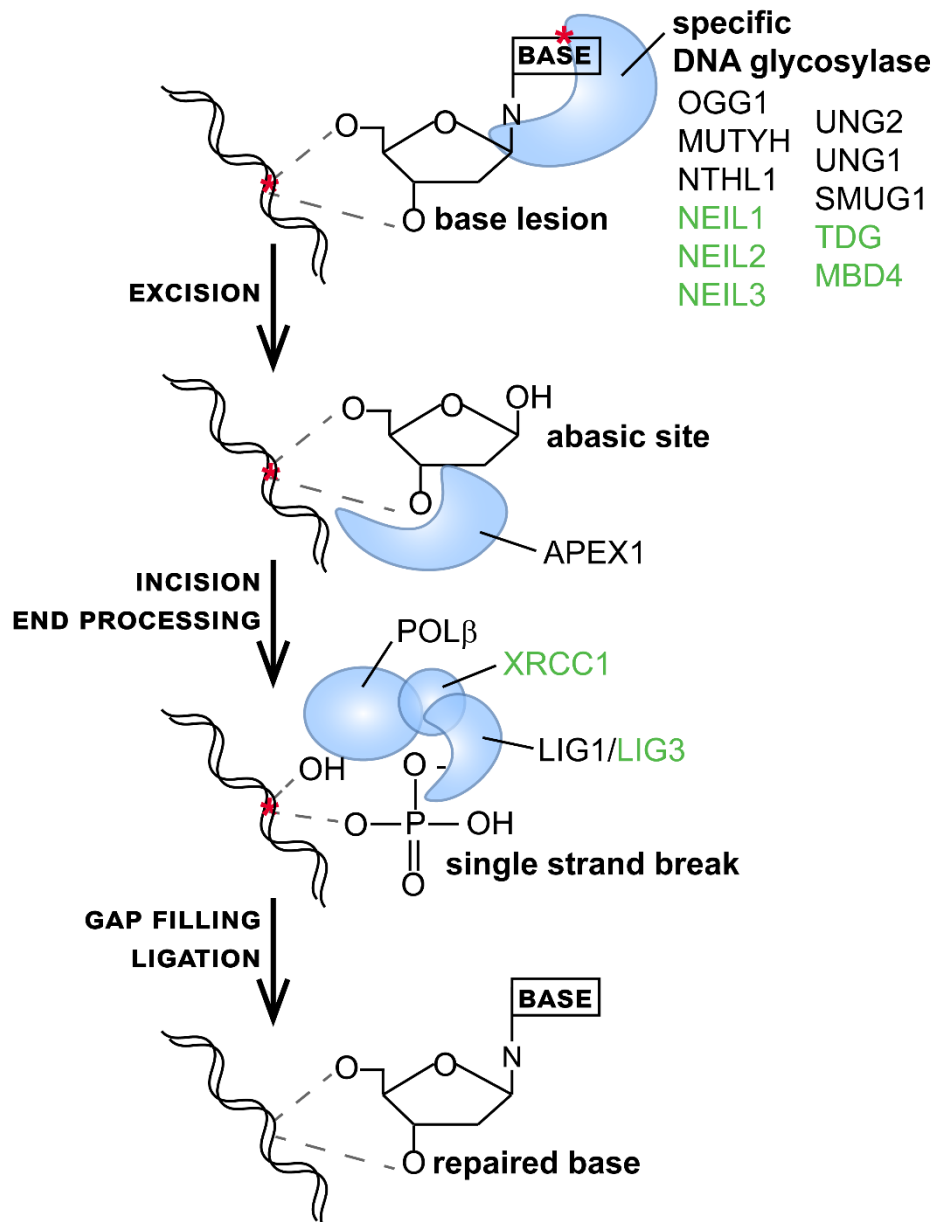
With a reporter gene approach, we here describe reactivation of gene expression in ESCs when plasmids harboring mC are oxidized by TET proteins. To unravel the impact of different DNA glycosylases on this observed gene reactivation, which is due to DNA demethylation, we performed the reporter gene assay with KO and rescue cell lines for different DNA glycosylases (Muller *et al.*, 2014). Two glycosylases in particular have been suggested to play a role in DNA demethylation: TDG and MBD4. Both enzymes do not excise hmC, but are catalytically active on T•G, U•G, and hmU•G mismatches which are generated by deamination of mC-G, C-G, and hmC-G base pairs, respectively (Hashimoto *et al.*, 2012). Additionally, TDG can remove fC and caC (He *et al.*, 2011; Maiti and Drohat, 2011).

According to our data, KO of *Mbd4* has no effect on reactivation of gene expression upon TET oxidation in ESCs, whereas *Tdg*-KO cells largely lose their ability to express the reporter gene. Furthermore, we show that the three NEIL glycosylases can partially rescue the loss of TDG in ESCs (Table 2) and that reactivation of reporter gene expression depends on the catalytic function and DNA binding ability of TDG (Muller *et al.*, 2014).

Taken together, our results lead to two key hypotheses. First, TET proteins and the BER pathway are functionally and physically linked and cooperatively cause active DNA demethylation by three

### 3. Discussion

steps: oxidation of mC by TETs, excision of fC and caC by TDG, and insertion of unmodified C by BER. Second, in the absence of TDG, NEIL glycosylases can excise fC and probably also caC, providing an alternative pathway for DNA demethylation.



**Figure 7: BER pathway and its interaction with TET proteins**

Schematic overview of the BER pathway. There are 11 known mammalian DNA glycosylases with different specificities towards certain types of base lesions. Proteins identified in this study as interaction partners of TET1, TET2, and TET3 are depicted in green (Krokan and Bjoras, 2013; Muller *et al.*, 2014).

The first hypothesis is strongly supported by a recent study (Weber *et al.*, 2016), which not only confirms the interaction between TET1 and TDG, but also reconstitutes the entire TET-BER pathway *in vitro*. Weber *et al* furthermore show that the catalytic domain of TET1 has a

stimulatory effect on TDG's base excision activity and that the BER machinery processes symmetrically modified DNA in a coordinated manner to avoid formation of double strand breaks (DSBs).

glycosylase	impaired gene reactivation	<i>Tdg</i> -KO rescue
TDG	yes	yes
MBD4	no	no
NEIL1	not tested	partial
NEIL2	not tested	partial
NEIL3	not tested	partial

**Table 2. Effect of DNA glycosylases on active DNA demethylation**

The five glycosylases interacting with TET proteins were screened for their impact on reactivation of a reporter gene in ESCs, when the plasmid carrying the reporter is previously oxidized by TETs. The second column indicates whether ESCs deficient for the respective glycosylase lose their ability to reactivate the reporter; the third column states whether expression of the glycosylase can compensate for loss of *Tdg* (Muller *et al.*, 2014).

In this thesis, TDG was also independently identified as a prominent reader protein for fC and caC (Spruijt *et al.*, 2013), in agreement with its excision activity on these bases. Furthermore, the glycosylases MPG and NEIL3 were described as binders of hmC in ESCs, whereas NEIL1 displayed affinity for all three oxidized cytosine variants: hmC, fC, and caC (Spruijt *et al.*, 2013). Recently, the role of NEIL glycosylases in fC and caC processing was studied in more detail by Schomacher *et al.* They describe that double knockdown (KD) of NEIL1 and NEIL2 strongly reduces fC and caC excision efficiency in HeLa cell extracts, comparable to KD of TDG. However, neither NEIL1 nor NEIL2 display fC or caC excision activity *in vitro*. The data by Schomacher *et al.* alternatively suggest a stimulatory effect of NEIL1 and NEIL2 on TDG activity. They propose a model where NEIL1 or NEIL2 displace TDG from the abasic site and act as AP (apurinic/apyrimidinic) lyases (Schomacher *et al.*, 2016). NEIL proteins are bifunctional glycosylases that not only possess base excision but also abasic site removal activity (Krokan and Bjoras, 2013). Schomacher *et al.* therefore state that the AP lyase activity rather than the glycosylase activity of NEILs is crucial for fC and caC processing. They additionally show a central role for NEIL2, TDG, and TET3 in neural crest formation in *Xenopus laevis* development, suggesting a cooperation between these three enzymes (Schomacher *et al.*, 2016).

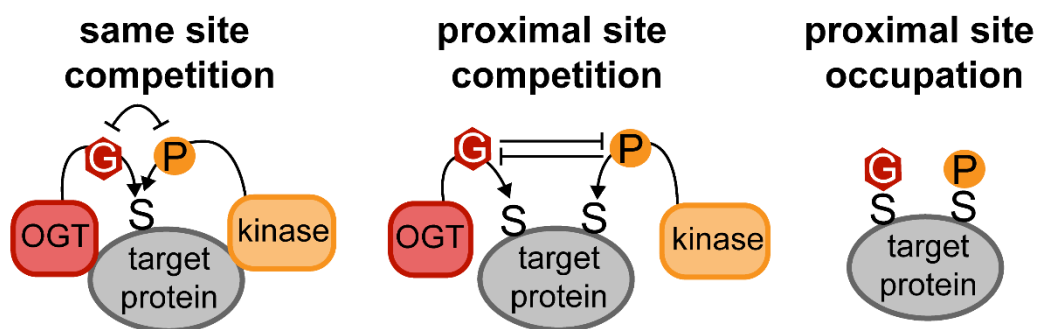
Our data, however, hint towards a direct role of all three NEIL proteins on excision of fC and/or caC in ESCs in the absence of TDG (Muller *et al.*, 2014). NEIL glycosylase activity on fC and caC might therefore depend on cellular co-factors that have not been identified to date and might also differ between species and developmental stages.

### 3.1.2 Phosphorylation and O-GlcNAcylation of TET proteins

In addition to proteins of the BER pathway, the O-GlcNAc transferase OGT co-precipitates with TET1, TET2, and TET3 expressed in HEK293T cells, as well as with endogenous TET1 and TET2

### 3. Discussion

in ESCs. We here describe that OGT causes *O*-GlcNAcylation of all three TET proteins and thereby reduces TET phosphorylation. Interestingly, this reduction is not due to competition of OGT and a yet unknown kinase for the same serine or threonine residue, since the majority of *O*-GlcNAcylation and phosphorylation sites are detected at distinct amino acids (Bauer *et al.*, 2015). Crosstalk between *O*-GlcNAcylation and phosphorylation has been described for several years; in fact, every protein of the approximately 1000 targets of OGT can also be phosphorylated. Three principle modes of this crosstalk are observed: first, same site competition, second, proximal site competition, and third, proximal site occupation (Figure 8) (Butkinaree *et al.*, 2010).



**Figure 8: *O*-GlcNAcylation-phosphorylation crosstalk**

Schematic depiction of the three modes of crosstalk between *O*-GlcNAcylation and phosphorylation. Same site competition: presence of one PTM on the residue inhibits addition of the other modification. Proximal site competition: presence of *O*-GlcNAcylation or phosphorylation inhibits modification of a proximal residue. Proximal site occupation: both types of modification co-reside on neighboring residues. It needs to be emphasized that *O*-GlcNAcylation dynamics are regulated by a single pair of enzymes, OGT and OGA, whereas there are more than 500 protein kinases and phosphatases (Butkinaree *et al.*, 2010).

Generally, competition between the two types of modification appears to be the more common form of crosstalk. For example, activation of protein kinase A or C results in decreased global *O*-GlcNAc levels, whereas inhibition of these kinases leads to broad upregulation of *O*-GlcNAcylation (Griffith and Schmitz, 1999). Furthermore, increase of *O*-GlcNAcylation results in mostly decreased phosphorylation, especially when there is an active turnover of phosphate groups (Wang *et al.*, 2008). However, the inverse relationship between phosphorylation and *O*-GlcNAcylation is not black or white. Inhibition of GSK3, for example, results in twice as many proteins with decreased rather than increased *O*-GlcNAc levels (Wang *et al.*, 2007). For TET proteins, we observe two types of phosphorylation: sites that show reduced phosphorylation occupancy upon increased *O*-GlcNAcylation and sites with constantly high phosphorylation occupancy, called “persistent phosphorylation” (Bauer *et al.*, 2015). This finding emphasizes the complex interplay between the two different PTMs and highlights the tight regulation of TET protein function.



The fact that TET proteins are *O*-GlcNAcylated by OGT (Ito *et al.*, 2014; Zhang *et al.*, 2014; Bauer *et al.*, 2015) immediately suggests that they can dynamically respond to the availability of nutrients and metabolites such as glucose, glutamine, acetyl-CoA, UTP, and ATP, since OGT activity depends on the hexosamine biosynthetic pathway. Indeed, Zhang *et al.* have shown that TET3 is exported from the nucleus when cells are cultured in high glucose medium (Zhang *et al.*, 2014).

The interplay between metabolism and epigenetics has interested scientists for decades since it represents an obvious means for a cell to respond to changing environmental conditions. The most direct effect of metabolism on epigenetics is by sensitivity towards cellular levels of co-factors and substrates for epigenetic modifiers. Table 3 provides an overview about the most important metabolites and vitamins consumed by epigenetic enzymes. ATP is not listed since intracellular ATP levels are usually not rate-limiting for kinase activity, however, histone phosphorylation can be indirectly influenced by ATP:AMP ratios (Lu and Thompson, 2012).

metabolite/vitamin	role	pathway/role	dependent enzymes
SAM	donor	one-carbon pathway	DNA and histone methyltransferases
methionine	precursor		
folate	precursor		
2-OG	co-substrate	TCA cycle	TET proteins and most histone lysine demethylases (Fe(II) and 2-OG dependent dioxygenases)
ascorbate	co-factor	reduction of Fe(IV)	
acetyl-CoA	donor	glycolysis, $\beta$ -oxidation of fatty acids	histone acetyltransferases
NAD <sup>+</sup>	co-factor	NAD <sup>+</sup> salvage pathway, oxidative phosphorylation	sirtuin family of histone deacetylases
tryptophan	precursor	de novo synthesis of NAD <sup>+</sup>	

**Table 3: Metabolites and vitamins that directly contribute to epigenetic regulation**

TCA: tricarboxylic acid, NAD<sup>+</sup>: nicotinamide adenine dinucleotide, NADH: reduced NAD<sup>+</sup> (Kanehisa and Goto, 2000; Lu and Thompson, 2012; Janke *et al.*, 2015).

Ascorbate (or vitamin C) is important for reduction of Fe(IV) to Fe(II) and therefore for the activity of Fe(II) and 2-OG dependent dioxygenases such as TET proteins (Janke *et al.*, 2015). Consequently, several groups have studied the influence of ascorbate on TET enzymatic activity (Blaschke *et al.*, 2013; Chen *et al.*, 2013; Dickson *et al.*, 2013; Yin *et al.*, 2013). Blaschke *et al.* report that addition of ascorbate to the medium of ESCs causes an increase in hmC levels and demethylation of mostly germ line genes, ultimately leading to a gene expression pattern that more closely resembles the ICM than culture without ascorbate (Blaschke *et al.*, 2013). Mechanistically, ascorbate seems to bind to the catalytic domain of TET proteins and directly enhances generation of all three mC oxidation products (Yin *et al.*, 2013).

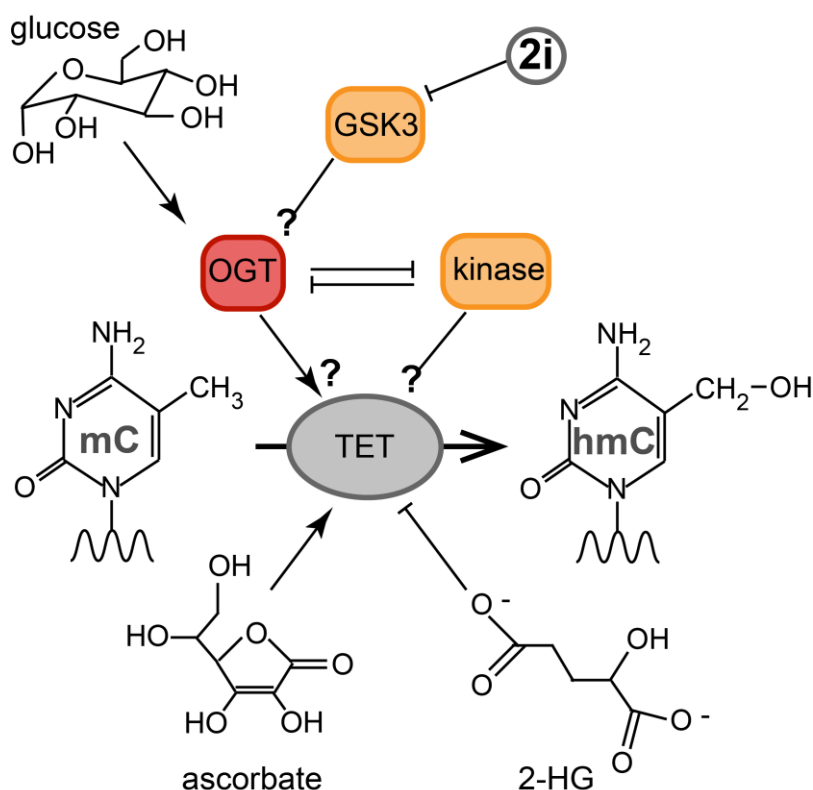
### 3. Discussion

Another direct link between TET activity and metabolism is represented by the oncometabolite 2-hydroxyglutarate (2-HG). 2-HG accumulates in glioblastoma and acute myeloid leukemia (AML) cells that carry a point mutation in the *IDH1* (isocitrate dehydrogenase) or *IDH2* gene. This mutation leads to a gain-of-function of the respective IDH enzyme (Parsons *et al.*, 2008; Dang *et al.*, 2009; Mardis *et al.*, 2009; Yan *et al.*, 2009). While wt IDH catalyzes the NADP<sup>+</sup> dependent oxidative decarboxylation of isocitrate to 2-OG, the oncogenic mutants further reduce 2-OG to 2-HG and, in contrast to wt IDH, can also carry out the reverse reaction from 2-OG to isocitrate. Thus, in addition to 2-HG accumulation, normal cellular levels of 2-OG and the NADP<sup>+</sup>/NADPH pool are also perturbed (Dang *et al.*, 2010). Interestingly, 2-HG acts as a competitive inhibitor of Fe(II) and 2-OG dependent dioxygenases such as TET proteins (Xu *et al.*, 2011). In a study with 30 patients with secondary AML, low levels of genomic hmC coincided with either a *TET2* mutation or an *IDH2* gain-of-function mutation in about 50 % of the cases (Konstandin *et al.*, 2011). This finding implies that both direct perturbation of TET protein sequence and inhibition of TETs by 2-HG represent alternative paths to lower hmC levels in secondary AML. However, other yet undiscovered mechanisms also seem to play an important role in regulating TET activity since in 50 % of the cases, low hmC levels could not be explained by the known mutations of TET and IDH enzymes.

PTMs of chromatin proteins add an additional layer of complexity to the interplay between metabolic and epigenetic pathways beyond the direct effects of co-factors. We here describe for the first time that TET proteins are not only *O*-GlcNAcylated but also phosphorylated (Bauer *et al.*, 2015). The biological function of the detected modification sites remains to be elucidated. *O*-GlcNAcylation apparently does not affect TET2 or TET3 enzymatic activity (Chen *et al.*, 2012; Deplus *et al.*, 2013), however, Vella *et al* report decreased hmC levels at certain genomic loci upon OGT depletion in ESCs (Vella *et al.*, 2013). Regarding our finding that TET *O*-GlcNAcylation reduces phosphorylation, which suggests a dynamic regulatory switch, and taking into consideration that TET activity is influenced by metabolism, nutrients, and vitamins as outlined above, cell culture conditions might be an important variable in studying TET protein regulation and also might explain discrepancies between different studies.

For example, standard cell culture media contain 25 mM glucose, a concentration comparable to the blood glucose level of leptin-deficient obese mice (Schwartz *et al.*, 1996) and therefore considerably higher than physiological glucose concentrations. Furthermore, perturbation of GSK3 signaling, a target of the “2i” inhibitors in ESC culture, has been shown to affect protein *O*-GlcNAcylation (Wang *et al.*, 2007). Finally, ascorbate, which is not a common medium additive, is likely to be crucial for proper TET function (Blaschke *et al.*, 2013). All these factors combined might influence PTM patterns as well as enzymatic activity of TET proteins in cell culture and

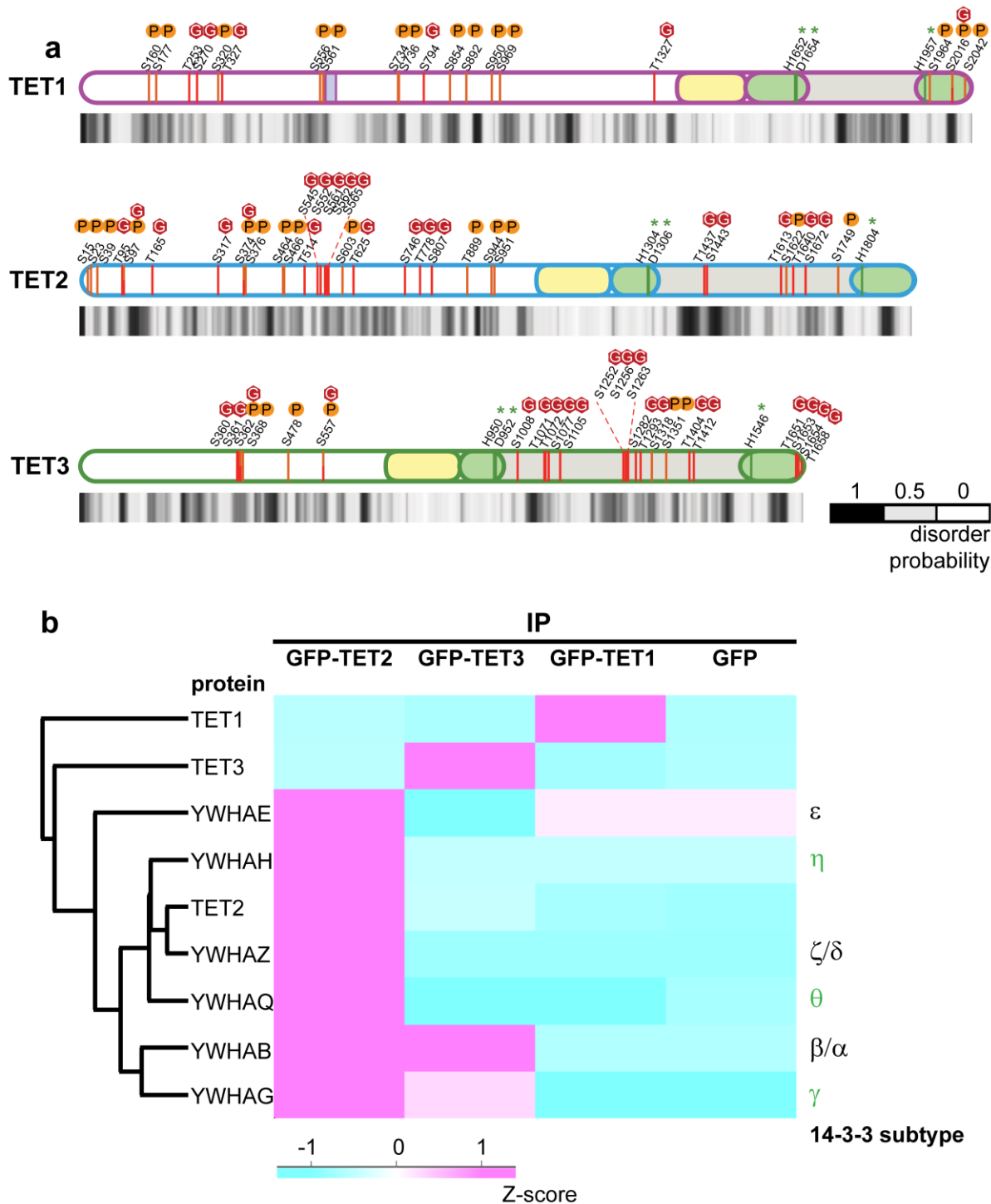
therefore perturb the picture of TET cellular function. Potential metabolic and nutritional regulators of TET proteins are summarized in Figure 9.



**Figure 9: Environmental conditions likely to affect TET protein regulation**

Ascorbate and 2-HG directly stimulate and inhibit TET activity, respectively (Xu *et al.*, 2011; Blaschke *et al.*, 2013). Protein O-GlcNAcylation by OGT depends on glucose levels and counteracts the phosphorylation activity of a yet unknown kinase (Butkinaree *et al.*, 2010; Harwood and Hanover, 2014; Bauer *et al.*, 2015). The 2i cocktail in ESC medium might also affect OGT activity via GSK3 inhibition (Wang *et al.*, 2007; Ying *et al.*, 2008).

The dynamic PTMs of TET proteins might also influence protein-protein interactions. Interestingly, several 14-3-3 type proteins, known for their ability to bind phosphorylated peptides, have been identified as interaction partners of TET2 (Deplus *et al.*, 2013). Considering the fact that we detect phosphorylation of TETs mostly at the low complexity regions (Figure 10a), the N-terminus and the insert, and that 14-3-3 proteins show high preference for intrinsically disordered proteins (Bustos and Iglesias, 2006; Uhart and Bustos, 2014), this suggests that 14-3-3 proteins are a dynamic reader module of TET proteins, that associate with the phosphorylated, disordered parts of the proteins and thus influence TET regulation.



**Figure 10: TET protein disorder and interaction with 14-3-3 proteins**  
a) PTMs of TET proteins as determined in (Bauer *et al.*, 2015) with prediction of protein disorder. The disorder probability for each residue is depicted by grayscale intensity; predictions were performed with PrDOS (Ishida and Kinoshita, 2007).  
b) TET proteins interact with 14-3-3 proteins. IP of GFP-TETs from HEK293T cells was followed by LC-MS/MS and results were analyzed with MaxQuant and Perseus (Cox and Mann, 2008) as described in (Bauer *et al.*, 2015). The heatmap shows Z-scored log<sub>2</sub> LFQ intensities (Cox *et al.*, 2014) for TET1, TET2, TET3, and the 14-3-3 proteins detected in the pull-down (unpublished results). On the left, protein names are shown based on the official gene names; the corresponding 14-3-3 nomenclature is depicted on the right. 14-3-3 subtypes that are also described by (Deplus *et al.*, 2013) as interaction partners of TET2 are written in green.

Consistent with the findings by Deplus *et al.*, we identify 14-3-3 proteins specifically interacting with TET2. We detect six out of seven 14-3-3 subtypes in the TET protein IPs, where only subtype  $\beta$  additionally interacts with TET3 (Figure 10b). It remains to be determined why TET2 is specifically bound by 14-3-3 proteins, since both phosphorylation pattern and disorder probability are very similar among TET1, TET2, and TET3, and would suggest equal affinity of 14-3-3 proteins towards all three TETs. In the case of TET proteins, low structural complexity and high phosphorylation levels seem to be necessary but not sufficient for interaction with 14-3-3. Also, none of the TET proteins contains one of the classic binding motifs of 14-3-3 proteins, RSXpSX<sup>3</sup>P or RXX<sup>1</sup>X<sup>2</sup>pSX<sup>3</sup>P (with X<sup>1</sup> optimally being an aromatic, X<sup>2</sup> a basic, and X<sup>3</sup> an L, E, A, or M residue) (Muslin *et al.*, 1996; Yaffe *et al.*, 1997). However, many 14-3-3 binding partners have been described which do not harbor the ideal interaction motifs (Bridges and Moorhead, 2004). Considering the proposed functions of 14-3-3 proteins (Bridges and Moorhead, 2004; Bustos and Iglesias, 2006), their interaction with TET2 could result in one or more of the following outcomes: first, binding of 14-3-3 might lead to a disorder-to-order transition at specific regions of TET2, second, important structural features of TET2 might be masked by 14-3-3 proteins, or third, 14-3-3 proteins might serve as interaction scaffolds due to their homo- and heterodimerization capability, and thus mediate interaction of TET2 with other proteins.

### 3.1.3 TET1 and chromatin modifiers

The first screens for TET protein interaction partners presented in this thesis were performed with GFP-tagged TETs expressed under the control of exogenous promoters. Interacting proteins were identified by IP and subsequent LC-MS/MS analysis (Muller *et al.*, 2014; Bauer *et al.*, 2015). When studying endogenous TET1 and TET2 in ESCs, this classic pull-down approach proved to be difficult due to low expression levels and protein instability in solution. For this reason, we applied the BioID method (Roux *et al.*, 2012), where a promiscuous biotin ligase is inserted into the native *Tet1* locus directly after the start codon (Mulholland *et al.*, 2015). Expression of the resulting fusion protein BirA\*-TET1 under the control of the endogenous Tet1 promoter leads to biotin-labeling of proteins within a 10 to 20 nm radius (Kim *et al.*, 2014) of TET1. Although not providing information about direct interaction partners, the BioID method is a useful tool to understand the functional protein environment of TET proteins.

Application of the BioID method led to identification of nine proteins within proximity of TET1 in ESCs, most of them linked to chromatin modifications (Mulholland *et al.*, 2015). Among those is SIN3A, a protein previously identified as a direct interaction partner of TET1, but not TET2 (Williams *et al.*, 2011; Deplus *et al.*, 2013). SIN3A acts as the central scaffold of the Sin3 co-repressor complex, coupling HDACs to transcriptional repressors (McDonel *et al.*, 2009). TET1 and SIN3A co-localize at TSS in ESCs, suggesting a role for TET1 in gene repression (Williams *et*

### 3. Discussion

*al.*, 2011). Another protein identified by TET1-BioID that is associated with transcriptional silencing is ZNF281, which is capable of recruiting the NuRD complex (Fidalgo *et al.*, 2012).

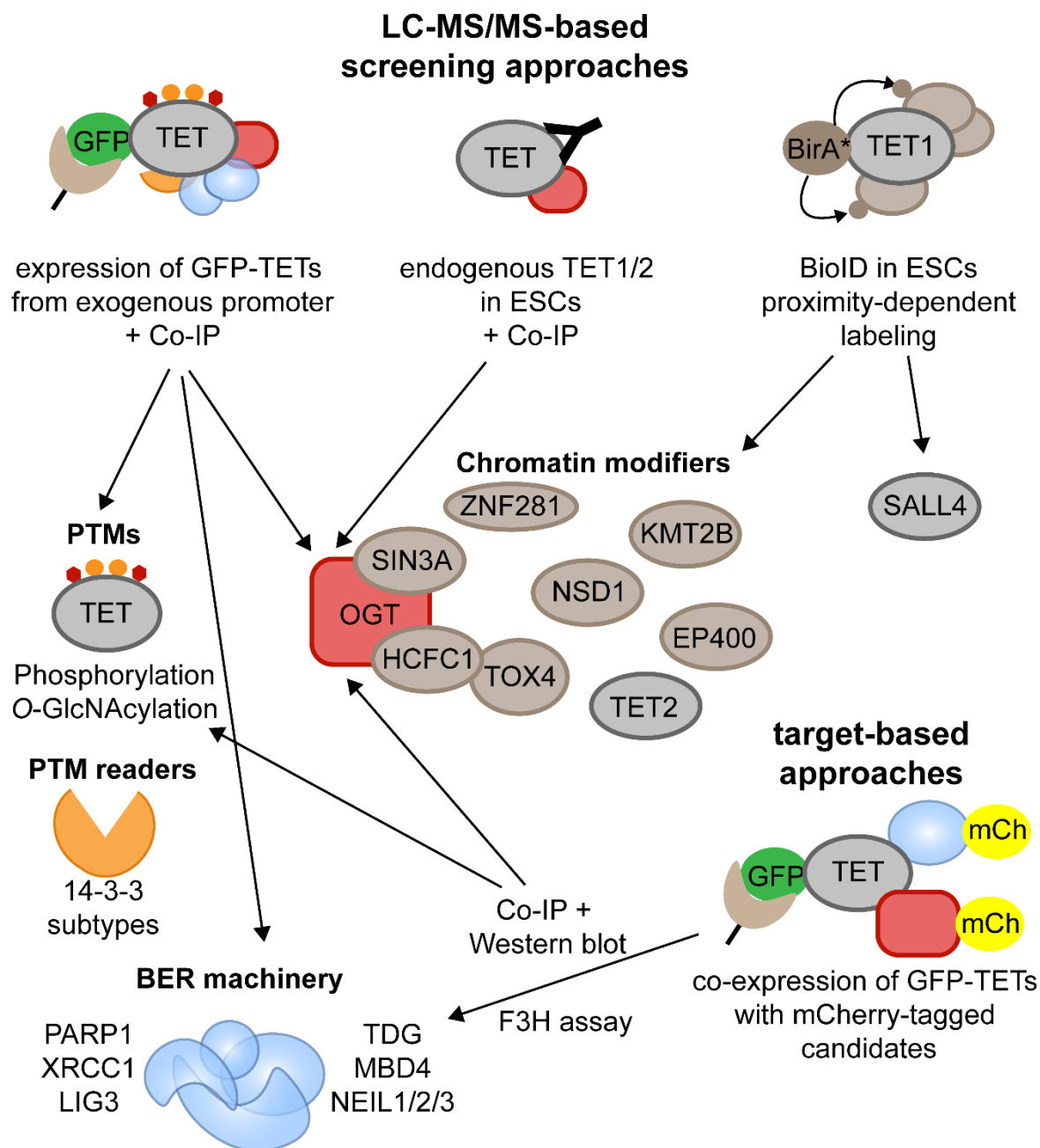
We furthermore detect HCFC1 (host cell factor 1) in proximity of TET1, a protein that also has also been suggested previously to interact with TET proteins (Cartron *et al.*, 2013; Deplus *et al.*, 2013). Deplus *et al.* describe that HCFC1 is *O*-GlcNAcylated in a TET2/3 and OGT dependent manner and that this modification is important for association of OGT with the SET1/COMPASS H3K4 methyltransferase complex. This finding links the combined activity of TET proteins and OGT to transcriptional activation. In light of the fact that TET1 is also associated with transcriptional repressors such as SIN3A, our findings confirm the suggested dual role of TET1 in both transcriptional activation and repression in ESCs (Wu *et al.*, 2011). Vella *et al.* have previously suggested complex formation between OGT, TET1, TET2, HCFC1, and SIN3A with OGT as the major component that mediates *O*-GlcNAcylation of the other complex members (Vella *et al.*, 2013). Interestingly, *O*-GlcNAcylation of HCFC1 not only depends on OGT, but also on GSK3 activity where inhibition of the kinase leads to a reduction of *O*-GlcNAcylation (Wang *et al.*, 2007). According to Wang *et al.*, HCFC1 is exported from the nucleus upon this reduction, therefore displaying the opposite behavior when compared to TET3 (Zhang *et al.*, 2014). In addition to HCFC1, we identify TOX4 by BioID, which also links TET proteins to the SET1 complex. TOX4 is a member of the PTW/PP1 phosphatase complex that interacts with the SET1 complex via WDR82 (Lee *et al.*, 2010).

EP400 (also referred to as mDomino) is also associated with TET1 in ESCs and links TET proteins to transcriptional activation since it is a component of the NuA4 histone acetyltransferase complex and additionally harbors chromatin remodeling activity (Doyon and Cote, 2004). Two histone methyltransferases, KMT2B and NSD1, are also detected by BioID in the proximity of TET1. KMT2B (also MLL2) is the catalytic member of the COMPASS complex that mediates H3K4me3 formation at bivalent promoters in ESCs (Hu, Garruss, *et al.*, 2013). NSD1, on the other hand, specifically methylates H3K36 and H4K20 (Rayasam *et al.*, 2003) and contains both activation and repression domains (Huang *et al.*, 1998).

Additionally, the BioID method identifies TET2 to be associated with TET1 (Mulholland *et al.*, 2015). At first sight, this is surprising since TET1 and TET2 have been assigned distinct roles in ESCs, with TET1 predominantly localizing to TSS and TET2 to gene bodies and enhancers (Williams *et al.*, 2011; Hon *et al.*, 2014; Huang *et al.*, 2014). However, a substantial number of TET1 and TET2 target genes overlap, and both TET1 and TET2 have been shown to interact with NANOG and contribute to regulation of pluripotency (Costa *et al.*, 2013; Huang *et al.*, 2014).

In summary, the majority of proteins that we identify as TET1-associated factors by BioID are involved in chromatin modification, namely histone acetylation, methylation, and deacetylation,

as well as methylcytosine oxidation. The only exception is SALL4, a central transcription factor which is required for ESC maintenance and development of ICM-derived cell lineages (Elling *et al.*, 2006). Our findings suggest a concerted action of several epigenetic complexes and pathways with TET1 in ESCs, possibly contributing to stem cell identity via SALL4.



**Figure 11: Regulation of TET proteins by PTMs and interacting proteins**

Overview of the methods that were applied in this thesis to analyze the pathways contributing to TET protein function (Muller *et al.*, 2014; Bauer *et al.*, 2015; Mulholland *et al.*, 2015). It is noteworthy that both F3H and BioID are *in vivo* methods, which provide an intact chromatin environment, whereas co-IP based approaches depend on *in vitro* stability of proteins and protein complexes.

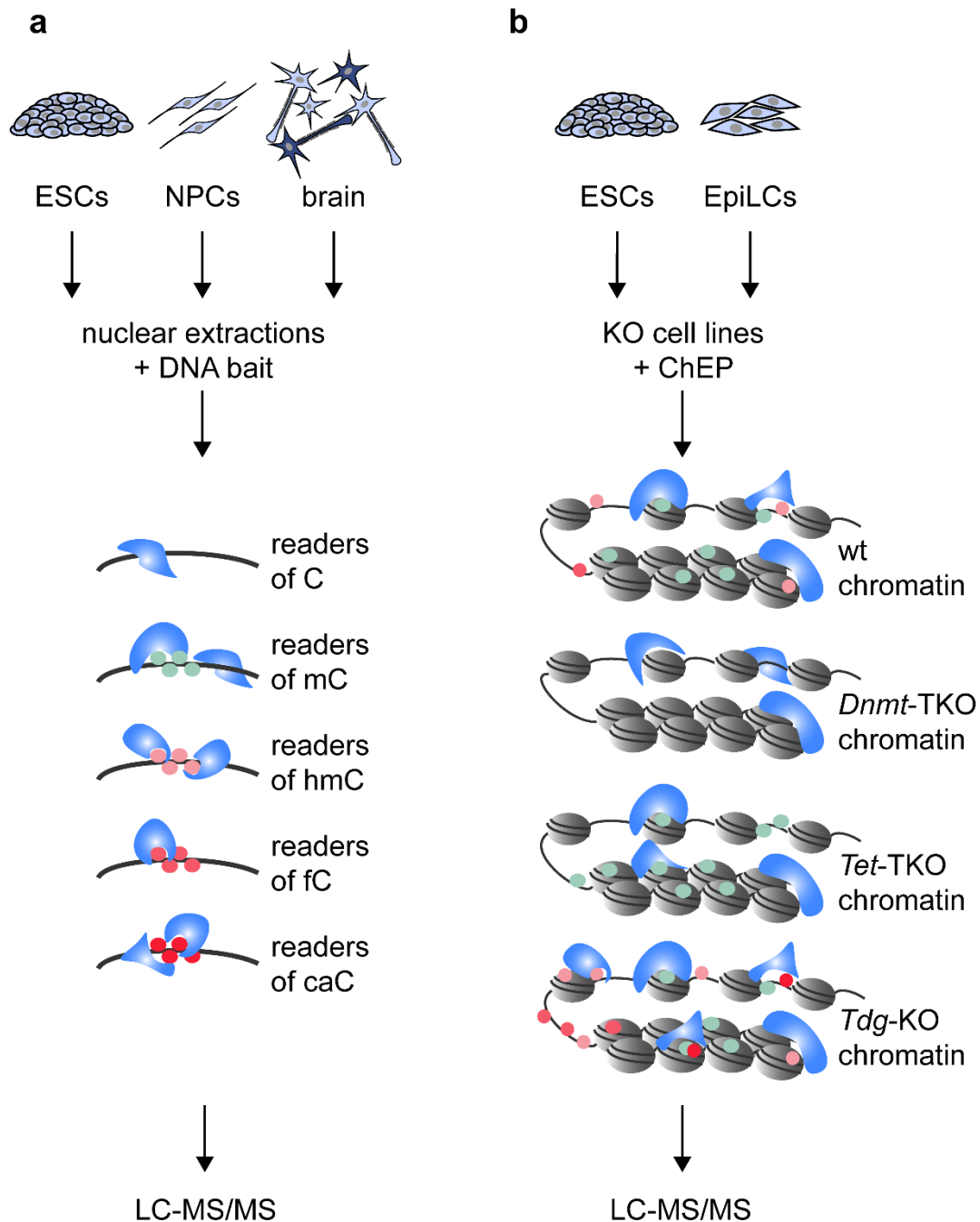


### 3. Discussion

All in all, we have characterized the regulation of TET proteins by interacting proteins and PTMs with a variety of different methods (Figure 11). We have applied LC-MS/MS-based screening approaches for *de novo* detection of functional interaction partners like BER proteins, OGT, 14-3-3 proteins, and chromatin modifiers and for identification and quantification of PTMs. We have confirmed and expanded these results by classic co-IP (co-immunoprecipitation) followed by Western blot detection and by the F3H (fluorescent-3-hybrid) assay, a method for analysis of protein-protein-interactions *in vivo* (Herce *et al.*, 2013). The obtained results are in agreement with the previous and current literature on TET proteins and on the identified interaction partners and contribute to the understanding of the fine-tuned regulation of TET proteins. We here describe three potential key functions of TET proteins: first, initiation of active DNA demethylation by association with DNA repair factors, second, dynamic response to a changing cellular environment via their connection to OGT, and third, regulation of both repressive and activating chromatin states.

#### 3.2 Interplay between cytosine modifications and chromatin proteins

The structural and functional association of genomic DNA with proteins defines chromatin states and is therefore essential for epigenetic regulation. In this thesis, we aim to understand the interplay between cytosine modifications, in particular the oxidative derivatives generated by TET proteins, and chromatin proteins during differentiation by two different approaches. The first approach is a bait-based pull-down method, where differentially modified DNA oligonucleotides are incubated with nuclear lysates of ESCs, NPCs, and adult murine brain tissue. Reader proteins bound to the DNA baits are identified and quantified by LC-MS/MS, either based on SILAC (stable isotope labeling by amino acids in cell culture) or LFQ (label-free quantification) (Spruijt *et al.*, 2013) (Figure 12a). The second approach makes use of a protocol for chromatin enrichment for proteomics (ChEP) (Kustatscher *et al.*, 2014) and compares the proteinaceous chromatin composition of ESCs that lack the major DNA modifiers (*Dnmt*-triple-knockout (TKO), *Tet*-TKO, and *Tdg*-KO) and therefore have an impaired DNA modification landscape. Furthermore, these ESC lines are differentiated to EpiLCs to compare epigenetic changes at this crucial developmental transition (Figure 12b).



**Figure 12: Approaches to elucidate the interplay between DNA modifications and chromatin proteins**

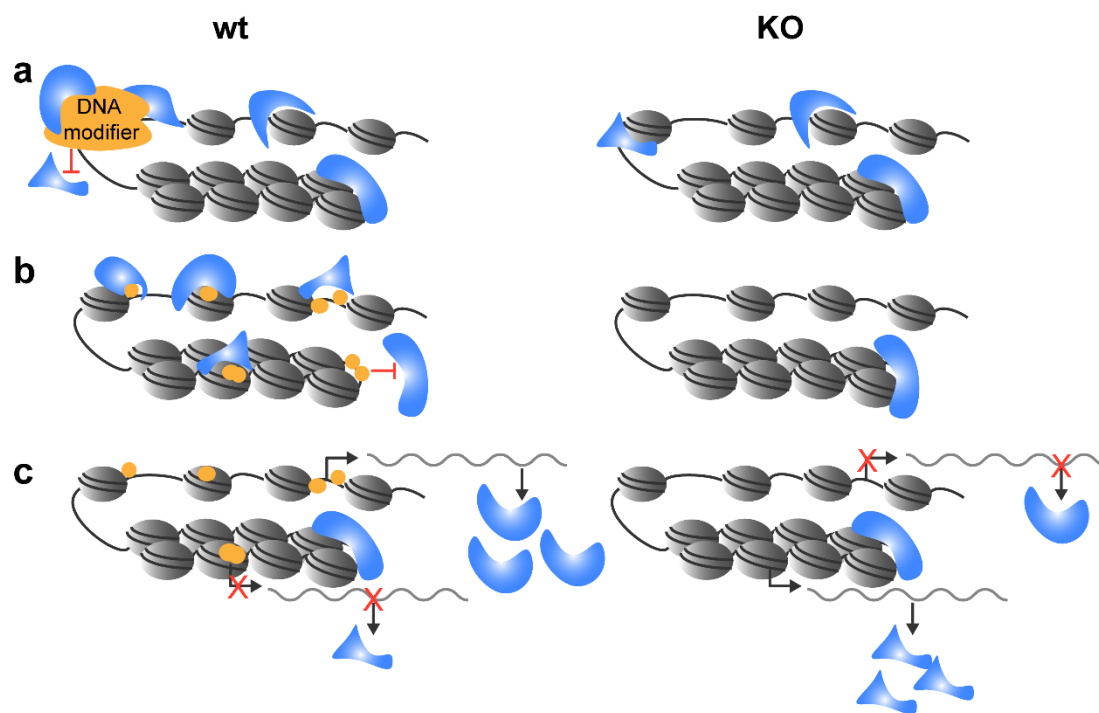
a) DNA pull-down approach as described in (Spruijt *et al.*, 2013). The DNA bait carries four consecutive modifications of the same type in a 5' - AAG ATG ATG ACG ACG ACG ACG ATG ATG - 3' sequence context with C being either C, mC, hmC, fC, or caC.

b) ChEP-based approach with isolated chromatin fractions from KO cell lines deficient for the major DNA modifiers, DNMTs, TETs, and TDG.

The two different experimental approaches have distinct advantages and disadvantages. In the DNA pull-down approach, defined substrates are provided as baits, leading to unambiguous results. By this method, direct readers of modified cytosines can be identified. However, unpackaged short DNA fragments resemble the natural target of chromatin proteins only poorly

### 3. Discussion

and might lead to false-negative results due to a lack of nucleosome structure and/or histone modifications. Additionally, proteins that strongly associate with genomic DNA are often hard to solubilize and might be underrepresented in the nuclear extracts. On the one hand, the ChEP approach overcomes these problems by considering the total chromatin fraction in different DNA modification backgrounds. On the other hand, the obtained results are more ambiguous since there are multiple explanations for the presence or absence of a protein in the respective chromatin fraction (Figure 13). For example, not only the abundance of the cytosine modification, but also of the corresponding DNA modifier is perturbed in the used KO cell lines, making it impossible to distinguish between proteins that are recruited by the modification or the modifier, respectively. Additionally, interfering with the total genomic DNA modification landscape by KO of *Dnmts*, *Tets*, or *Tdg* might epigenetically change gene expression and therefore protein levels and chromatin abundance. Our currently ongoing RNA-sequencing experiments of wt ESCs, EpiLCs, and the KO cell lines will help to understand the contribution of epigenetically induced perturbations of gene expression on chromatin composition.



**Figure 13: Explanations for differential chromatin composition in cells deficient for DNA modifiers**

Schematic illustration of the possible mechanisms that contribute to differential chromatin abundance of proteins in the ChEP-based approach. The situation in wt and KO cells is shown on the left and right, respectively.

a) The chromatin proteins are directly recruited or repelled by the DNA-modifying enzyme, either DNMTs, TETs, or TDG.

b) The DNA modifications (yellow circles) promote or inhibit binding of proteins to chromatin.

c) The DNA modifications influence gene transcription and thus, overall protein abundance.

Combination of the DNA pull-down results and the ChEP-based data provides valuable information about the interplay between DNA modifications and chromatin proteins. For example, proteins of the MBD (methyl CpG binding domain) family have been the first described specific binders of mC and are a classic example of readers of methylated DNA (Hendrich and Bird, 1998). There are five family members in mammals, MECP2, MBD1, MBD2, MBD3, and MBD4 (Fatemi and Wade, 2006), which all show higher affinity for mC than C with the exception of MBD3 (Fraga *et al.*, 2003) (Figure 14a). We here specifically pull down the MBD proteins MBD1 and MBD4 by mC-containing DNA and MECP2 by both mC and hmC in ESCs (Spruijt *et al.*, 2013). In contrast, we do not detect significantly different MBD levels in chromatin of *Dnmt*-TKO ESCs compared to wt. This is surprising since local genomic enrichment of MBD1, 2, 4, and MECP2 has been shown to correlate with mCpG density, which seems to be the major determinant of MBD localization (Baubec *et al.*, 2013). However, MECP2 does not lose its association with constitutive heterochromatin in *Dnmt*-TKO ESCs (Baubec *et al.*, 2013), suggesting additional factors that determine its chromatin affinity and might contribute to MECP2's persistent binding to the chromatin fraction of *Dnmt*-TKOs. Unfortunately, MBD1 and MBD4 are not covered by our ChEP-based dataset, allowing no conclusion about their overall chromatin abundance in the *Dnmt*-TKO background.

Interestingly, we find MECP2 enriched in chromatin of *Tdg*-KO ESCs, along with other heterochromatin proteins like H1 isoforms, TRIM28 (also known as KAP1), SIN3A, CBX3, and HP1BP3. A recent study on the effects of cytosine modifications on DNA flexibility attributes increased mechanical stability to nucleosomes containing fC (Ngo *et al.*, 2016). Since enrichment of fC is rather strong in *Tdg*-KO ESCs, this enhanced stability might contribute to recruitment of heterochromatin factors. Additionally, loss of TDG leads to aberrant DNA methylation at CGIs and an increase of repressive histone marks such as H3K9me3 and H3K27me3 at certain promoters in MEFs (mouse embryonic fibroblasts) (Cortazar *et al.*, 2011), supporting our finding of accumulation of heterochromatin-associated proteins. However, Cortazar *et al.* conclude that TDG induces epigenetic aberrations mostly in differentiated and not in pluripotent cells, which is in contrast to our results. It is noteworthy that the chromatin fractions analyzed in this study were obtained from ESCs cultured under naïve pluripotent conditions in the presence of 2i, whereas the data of Cortazar *et al.* were generated under serum/LIF conditions, which more resemble the epiblast than the ICM state (Cortazar *et al.*, 2011). These alternative culture conditions might lead to epigenetic changes, especially considering the large effects of the presence or absence of 2i on the DNA modification landscape (Ficz *et al.*, 2013).

While the DNA pull-down and the ChEP-based approach yield different results concerning MBD proteins, both methods lead to identification of KLF4 as a protein whose affinity for DNA depends on DNA methylation (Spruijt *et al.*, 2013). Additionally, chromatin enrichment of *Dnmt*-TKOs

### 3. Discussion

shows a depletion of OCT4, which is another of the four Yamanaka reprogramming factors KLF4, OCT4, SOX2, and C-MYC (Takahashi and Yamanaka, 2006). It seems plausible that the proteins required for reprogramming of differentiated to pluripotent cells display a preference for mC. The promoters of core pluripotency genes like *Nanog* are methylated in differentiated tissues and need to be reactivated during reprogramming (Meissner, 2010). Binding of KLF4 or OCT4 to these methylated areas might induce gene reactivation and thereby trigger the reprogramming process.

Both chromatin of *Tdg*-KO cells and fC pull-down strongly enrich for proteins associated with the GO terms “DNA repair” and “Response to DNA damage stimulus” (Spruijt *et al.*, 2013). Abundance of fC seems to recruit a variety of different repair pathways in ESCs, including BER, nucleotide excision repair, mismatch repair, and DSB repair. Additionally, the methylpurine DNA glycosylase (MPG) was identified as a specific reader for hmC in ESCs (Spruijt *et al.*, 2013) and is consistently depleted from *Tet*-TKO chromatin. In contrast, other studies report MPG as a reader for fC although without excision activity (Iurlaro *et al.*, 2013). These discrepancies highlight the complexity of the interplay between DNA modifications and their associated proteins. Several factors are likely to influence differential binding of a certain protein to modified DNA and shall be discussed below.

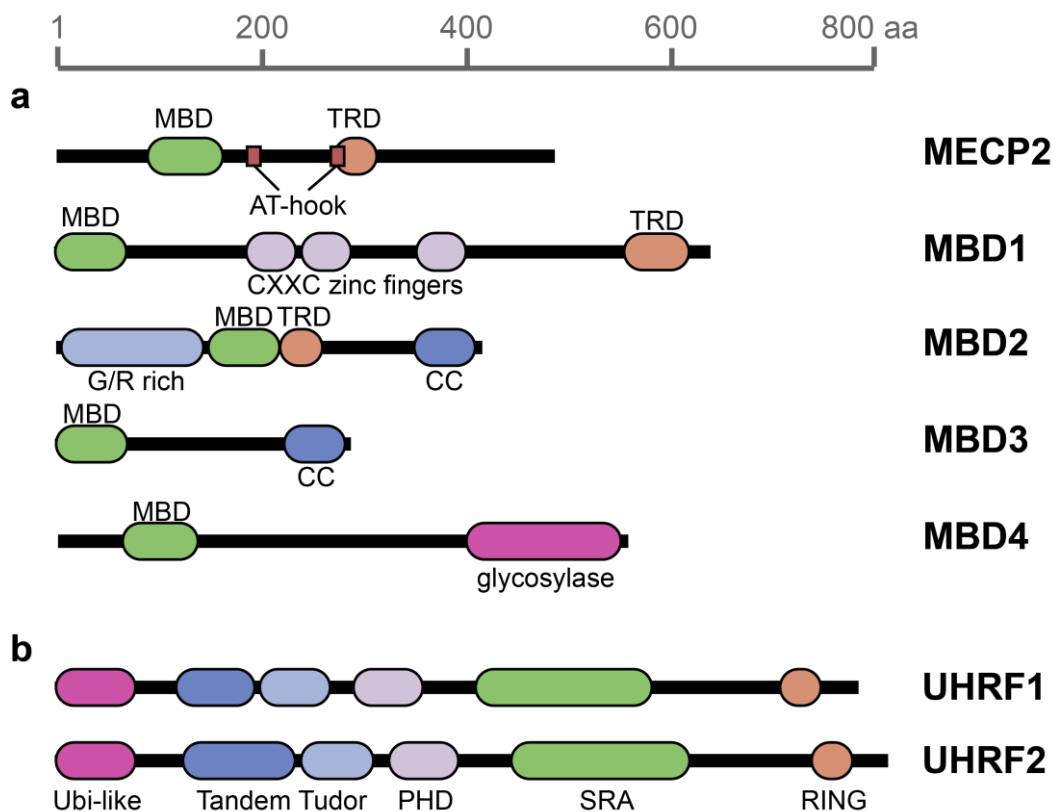
First, the sequence context of the DNA might play a role, especially with regards to transcription factors. In this study, we used a random 27 bp DNA sequence with four cytosine modification sites in an ACG context (Spruijt *et al.*, 2013). Iurlaro *et al.* generated modified DNA baits by PCR, using approximately 250 bp sequences of the *Fgf15* and the *Pax6* gene promoters, the first harboring a CGI. They e.g. detect an equally strong preference of UHRF1 for mC with both DNA baits, but binding of UHRF1 to hmC-containing DNA is less pronounced with the *Fgf15* DNA compared to *Pax6* DNA (Iurlaro *et al.*, 2013). With our DNA bait, UHRF1 displays equal affinity for both mC and hmC as has been shown by *in vitro* data before (Frauer *et al.*, 2011). Interestingly, the UHRF1 paralog UHRF2 (Figure 14b) has been detected as a reader protein for hmC specifically in NPCs. The cellular functions of UHRF2 are largely unknown, however, our data suggest an impact on TET oxidation activity (Spruijt *et al.*, 2013).

Second, the chromatin context, e.g. nucleosome occupancy and histone modifications, will influence binding. In the chromatin fractions of our ChEP-based study, UHRF1 abundances are comparable across all samples, suggesting that chromatinization of DNA counteracts the recruiting effects of DNA modifications. Consistently, intranuclear dynamics of UHRF1 are not affected by loss of DNA methylation and its DNA binding SRA domain is not sufficient to recruit UHRF1 to constitutive heterochromatin (Rottach *et al.*, 2010). Interestingly, an allosteric regulatory mechanism for UHRF1 was described recently. Via its PHD (plant homeodomain), UHRF1 binds to unmodified H3 tails, but in the presence of phosphatidylinositol-5-phosphate, a

conformational switch induces binding of H3K9me3 via the TTD (tandem tudor domain) (Gelato *et al.*, 2014) (Figure 14b).

Third, density or spacing of the DNA modifications could also influence protein binding. For example, Baubec *et al.* identify methylation density as the major determinant for genomic localization of MBD proteins, suggesting that both abundance and methylation status of CpG sites are important for MBD targeting to DNA (Baubec *et al.*, 2013).

Finally, it cannot be excluded that there are combinatorial effects of different cytosine modifications, although this aspect has not been investigated to date.



**Figure 14: Examples of DNA modification readers**

a) Scaled domain model of MBD proteins (UniProt-IDs Q9Z2D6, Q9Z2E2, Q9Z2E1, Q9Z2D8, Q9Z2D7). MBD: methyl CpG binding domain, TRD: transcriptional repression domain, CC: coiled-coil (Fatemi and Wade, 2006; Baubec *et al.*, 2013).

b) Schematic depiction of UHRF1 and UHRF2 (UniProt-IDs Q8VDF2, Q7TMI3). Ubi-like: ubiquitin-like domain, PHD: plant homeodomain, SRA: SET and RING associated, RING: E3 ubiquitin ligase domain.

While the crosstalk between DNA modifications and reader proteins is already complex in ESCs, it also dynamically changes during differentiation. The overlap of readers for mC and hmC in ESCs, NPCs, and brain is rather low (Spruijt *et al.*, 2013), suggesting that cytosine modifications play distinct roles in different cell types. In line with this observation, the chromatin fractions of the used KO cell lines greatly differ between ESCs and EpiLCs, implicating that differentiation state is

### 3. Discussion

a strong determinant of chromatin composition and that changes in the DNA modification landscape have different impacts in different cell types.

When comparing NPC and EpiLC differentiation, there is a striking difference in the behavior of chromatin proteins. The number of identified specific reader proteins for C, mC or hmC is lower in ESCs than in NPCs (69 versus 190 enriched interactors, respectively) (Spruijt *et al.*, 2013). Comparing the ChEP fractions of ESCs to EpiLCs, however, this relation is reversed and a higher number of proteins show significantly different chromatin association upon DNA modification depletion in ESCs than in EpiLCs. On the one hand, it cannot be excluded that these differences are due to intrinsic methodological variations between the DNA pull-down and the ChEP-based approach. On the other hand, it is surprising that loss of DNA methylation in the *Dnmt*-TKO background has so little effect in EpiLCs (134 chromatin proteins with differential abundance in ESCs and 20 differential proteins in EpiLCs), considering the strong upregulation of genomic cytosine methylation at this stage (Borgel *et al.*, 2010). In combination, these data suggest that DNA modifications have distinct roles and relevance in different cell types and developmental stages, where the EpiLC state is less affected by DNA modification patterns. This leads to the conclusion that cytosine modifications strongly influence chromatin composition in naïve pluripotent cells and more terminally differentiated cell types like NPCs, a finding also supported by gene expression and histone modification data in *Tdg*-KO cells (Cortazar *et al.*, 2011). It is important to note that cells resembling the "primed" epiblast state are considered metastable (Hayashi *et al.*, 2008; Hackett *et al.*, 2013) and therefore display greater heterogeneity. This might provide an explanation why EpiLCs or ESCs in serum/LIF are less susceptible to perturbations of the epigenetic environment.

In summary, we aimed to dissect the contribution of DNA modifications to epigenetic states by identification of direct readers for the five cytosine specimen C, mC, hmC, fC, and caC, and by analysis of the chromatin composition in cell lines deficient for important DNA modifiers like DNMTs, TETs, or TDG. Some proteins like MECP2 or UHRF1 display a strong intrinsic preference for certain DNA modifications, but are not affected by loss of these modification in an intact chromatin context, suggesting additional recruiting mechanisms. Other proteins like KLF4 or DNA repair proteins depend on the DNA modification status, regardless if the DNA substrate is chromatinized or not. We furthermore show that cytosine modification readers change during differentiation of ESCs towards the neuronal lineage and that chromatin in ground state pluripotency is more sensitive to changes in the DNA modification landscape than in the EpiLC state.



## References

- Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., Encheva, V., Kudo, N.R., Koseki, H., Sasaki, H., and Hajkova, P. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat Cell Biol* 18, 225-233.
- Argentaro, A., Yang, J.C., Chapman, L., Kowalczyk, M.S., Gibbons, R.J., Higgs, D.R., Neuhaus, D., and Rhodes, D. (2007). Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proc Natl Acad Sci U S A* 104, 11939-11944.
- Arney, K.L., and Fisher, A.G. (2004). Epigenetic aspects of differentiation. *J Cell Sci* 117, 4355-4363.
- Auclair, G., Guibert, S., Bender, A., and Weber, M. (2014). Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol* 15, 545.
- Auclair, G., and Weber, M. (2012). Mechanisms of DNA methylation and demethylation in mammals. *Biochimie* 94, 2202-2211.
- Avery, O.T., Macleod, C.M., and McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from *Pneumococcus* Type Iii. *J Exp Med* 79, 137-158.
- Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merckenschlager, M., and Fisher, A.G. (2006). Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* 8, 532-538.
- Baubec, T., Ivanek, R., Lienert, F., and Schubeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* 153, 480-492.
- Baubec, T., and Schubeler, D. (2014). Genomic patterns and context specific interpretation of DNA methylation. *Curr Opin Genet Dev* 25, 85-92.
- Bauer, C., Gobel, K., Nagaraj, N., Colantuoni, C., Wang, M., Muller, U., Kremmer, E., Rottach, A., and Leonhardt, H. (2015). Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J Biol Chem* 290, 4801-4812.
- Bedford, M.T. (2007). Arginine methylation at a glance. *J Cell Sci* 120, 4243-4246.
- Beltrao, P., Bork, P., Krogan, N.J., and van Noort, V. (2013). Evolution and functional cross-talk of protein post-translational modifications. *Mol Syst Biol* 9, 714.
- Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol* 203, 971-983.
- Bestor, T.H. (2000). The DNA methyltransferases of mammals. *Hum Mol Genet* 9, 2395-2402.
- Bhaumik, S.R., Smith, E., and Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol* 14, 1008-1016.
- Bibel, M., Richter, J., Lacroix, E., and Barde, Y.A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat Protoc* 2, 1034-1043.
- Bibel, M., Richter, J., Schrenk, K., Tucker, K.L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat Neurosci* 7, 1003-1009.
- Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martinez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., Lorincz, M.C., and Ramalho-Santos, M. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* 500, 222-226.
- Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 42, 1093-1100.
- Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760-1764.

## References

- Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science* 294, 2536-2539.
- Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* 309, 255-256.
- Bridges, D., and Moorhead, G.B. (2004). 14-3-3 proteins: a number of functions for a numbered protein. *Sci STKE* 2004, re10.
- Bustos, D.M., and Iglesias, A.A. (2006). Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins. *Proteins* 63, 35-42.
- Butkinaree, C., Park, K., and Hart, G.W. (2010). O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim Biophys Acta* 1800, 96-106.
- Callebaut, I., Courvalin, J.C., and Mornon, J.P. (1999). The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. *FEBS Lett* 446, 189-193.
- Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* 14, 155-164.
- Cardoso, M.C., and Leonhardt, H. (1999). DNA methyltransferase is actively retained in the cytoplasm during early development. *J Cell Biol* 147, 25-32.
- Cartron, P.F., Nadaradjane, A., Lepape, F., Lalier, L., Gardie, B., and Vallette, F.M. (2013). Identification of TET1 Partners That Control Its DNA-Demethylating Function. *Genes Cancer* 4, 235-241.
- Chedin, F., Lieber, M.R., and Hsieh, C.L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc Natl Acad Sci U S A* 99, 16916-16921.
- Chen, J., Guo, L., Zhang, L., Wu, H., Yang, J., Liu, H., Wang, X., Hu, X., Gu, T., Zhou, Z., Liu, J., Liu, J., Wu, H., Mao, S.Q., Mo, K., Li, Y., Lai, K., Qi, J., Yao, H., Pan, G., *et al.* (2013). Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat Genet* 45, 1504-1509.
- Chen, L., MacMillan, A.M., Chang, W., Ezaz-Nikpay, K., Lane, W.S., and Verdine, G.L. (1991). Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry* 30, 11018-11025.
- Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2012). TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 493, 561-564.
- Chen, T., Tsujimoto, N., and Li, E. (2004). The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Mol Cell Biol* 24, 9048-9058.
- Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* 277, 1996-2000.
- Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., Steinacher, R., Jiricny, J., Bird, A., and Schar, P. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* 470, 419-423.
- Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., Das, S., Levasseur, D.N., Li, Z., Xu, M., Reik, W., Silva, J.C., and Wang, J. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* 495, 370-374.
- Cox, J., Hein, M.Y., Lubner, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13, 2513-2526.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26, 1367-1372.
- Dang, L., Jin, S., and Su, S.M. (2010). IDH mutations in glioma and acute myeloid leukemia. *Trends Mol Med* 16, 387-397.
- Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., Marks, K.M., Prins, R.M., Ward, P.S., Yen, K.E., Liao, L.M., Rabinowitz, J.D., Cantley, L.C.,

- Thompson, C.B., Vander Heiden, M.G., and Su, S.M. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739-744.
- Dean, W., Santos, F., Stojkovic, M., Zakhartchenko, V., Walter, J., Wolf, E., and Reik, W. (2001). Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proc Natl Acad Sci U S A* 98, 13734-13738.
- Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev* 25, 1010-1022.
- Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Mendez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., Shih, A.H., Levine, R.L., Bernard, O., Mercher, T., Solary, E., Urh, M., Daniels, D.L., and Fuks, F. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J* 32, 645-655.
- Dickson, K.M., Gustafson, C.B., Young, J.I., Zuchner, S., and Wang, G. (2013). Ascorbate-induced generation of 5-hydroxymethylcytosine is unaffected by varying levels of iron and 2-oxoglutarate. *Biochem Biophys Res Commun* 439, 522-527.
- Doyon, Y., and Cote, J. (2004). The highly conserved and multifunctional NuA4 HAT complex. *Curr Opin Genet Dev* 14, 147-154.
- Du, Z., Song, J., Wang, Y., Zhao, Y., Guda, K., Yang, S., Kao, H.Y., Xu, Y., Willis, J., Markowitz, S.D., Sedwick, D., Ewing, R.M., and Wang, Z. (2010). DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. *Sci Signal* 3, ra80.
- Elling, U., Klasen, C., Eisenberger, T., Anlag, K., and Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. *Proc Natl Acad Sci U S A* 103, 16319-16324.
- Endicott, J.A., Noble, M.E., and Johnson, L.N. (2012). The structural basis for control of eukaryotic protein kinases. *Annu Rev Biochem* 81, 587-613.
- Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultschi, A.I., Wutz, A., and Bickmore, W.A. (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell* 38, 452-464.
- Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154-156.
- Ewing, B., and Green, P. (2000). Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet* 25, 232-234.
- Fatemi, M., and Wade, P.A. (2006). MBD family proteins: reading the epigenetic code. *J Cell Sci* 119, 3033-3037.
- Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J., Cook, S.J., Andrews, S., Branco, M.R., and Reik, W. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* 13, 351-359.
- Fidalgo, M., Faiola, F., Pereira, C.F., Ding, J., Saunders, A., Gingold, J., Schaniel, C., Lemischka, I.R., Silva, J.C., and Wang, J. (2012). Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. *Proc Natl Acad Sci U S A* 109, 16202-16207.
- Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* 2, 160-169.
- Fraga, M.F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P.A., and Esteller, M. (2003). The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res* 31, 1765-1774.
- Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M.C., Antes, I., and Leonhardt, H. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One* 6, e21306.
- Fu, J., Bian, M., Jiang, Q., and Zhang, C. (2007). Roles of Aurora kinases in mitosis and tumorigenesis. *Mol Cancer Res* 5, 1-10.
- Gambetta, M.C., and Muller, J. (2015). A critical perspective of the diverse roles of O-GlcNAc transferase in chromatin. *Chromosoma* 124, 429-442.

## References

- Gao, Y., Wells, L., Comer, F.I., Parker, G.J., and Hart, G.W. (2001). Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *J Biol Chem* 276, 9838-9845.
- Gelato, K.A., Tauber, M., Ong, M.S., Winter, S., Hiragami-Hamada, K., Sindlinger, J., Lemak, A., Bultsma, Y., Houliston, S., Schwarzer, D., Divecha, N., Arrowsmith, C.H., and Fischle, W. (2014). Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate. *Mol Cell* 54, 905-919.
- Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74, 481-514.
- Grewal, S.I., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. *Science* 301, 798-802.
- Griffith, L.S., and Schmitz, B. (1999). O-linked N-acetylglucosamine levels in cerebellar neurons respond reciprocally to perturbations of phosphorylation. *Eur J Biochem* 262, 824-831.
- Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G., Iqbal, K., Shi, Y.G., Deng, Z., Szabo, P.E., Pfeifer, G.P., Li, J., and Xu, G.L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606-610.
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145, 423-434.
- Gurley, L.R., D'Anna, J.A., Barham, S.S., Deaven, L.L., and Tobey, R.A. (1978). Histone phosphorylation and chromatin structure during mitosis in Chinese hamster cells. *Eur J Biochem* 84, 1-15.
- Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., Benedetti, R., Altucci, L., Jansen, J.H., Walter, J., Gut, I.G., Marks, H., and Stunnenberg, H.G. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* 13, 360-369.
- Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., and Surani, M.A. (2013). Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. *Stem Cell Reports* 1, 518-531.
- Hanover, J.A., Yu, S., Lubas, W.B., Shin, S.H., Ragano-Caracciola, M., Kochran, J., and Love, D.C. (2003). Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Arch Biochem Biophys* 409, 287-297.
- Harwood, K.R., and Hanover, J.A. (2014). Nutrient-driven O-GlcNAc cycling - think globally but act locally. *J Cell Sci* 127, 1857-1867.
- Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res* 40, 4841-4849.
- Hayashi, K., Lopes, S.M., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* 3, 391-401.
- Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* 146, 519-532.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.X., Zhang, K., He, C., and Xu, G.L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303-1307.
- Hemberger, M., Dean, W., and Reik, W. (2009). Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat Rev Mol Cell Biol* 10, 526-537.
- Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* 18, 6538-6547.
- Herce, H.D., Deng, W., Helma, J., Leonhardt, H., and Cardoso, M.C. (2013). Visualization and targeted disruption of protein interactions in living cells. *Nat Commun* 4, 2660.

- Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* 279, 48350-48359.
- Holliday, R. (2014). Epigenetics: A Historical Overview. *Epigenetics* 1, 76-80.
- Hon, G.C., Song, C.X., Du, T., Jin, F., Selvaraj, S., Lee, A.Y., Yen, C.A., Ye, Z., Mao, S.Q., Wang, B.A., Kuan, S., Edsall, L.E., Zhao, B.S., Xu, G.L., He, C., and Ren, B. (2014). 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell* 56, 286-297.
- Hu, D., Garruss, A.S., Gao, X., Morgan, M.A., Cook, M., Smith, E.R., and Shilatifard, A. (2013). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat Struct Mol Biol* 20, 1093-1097.
- Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y.G., Zhu, J., Wang, P., and Xu, Y. (2013). Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* 155, 1545-1555.
- Huang, N., vom Baur, E., Garnier, J.M., Lerouge, T., Vonesch, J.L., Lutz, Y., Chambon, P., and Losson, R. (1998). Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *EMBO J* 17, 3398-3412.
- Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W.A., Kang, J., Zepeda-Martinez, J.A., Pape, U.J., Jacobsen, S.E., Peters, B., and Rao, A. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* 111, 1361-1366.
- Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villén, J., Haas, W., Sowa, M.E., and Gygi, S.P. (2010). A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. *Cell* 143, 1174-1189.
- Illingworth, R.S., Moffat, M., Mann, A.R., Read, D., Hunter, C.J., Pradeepa, M.M., Adams, I.R., and Bickmore, W.A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. *Genes Dev* 29, 1897-1902.
- Inoue, A., Shen, L., Matoba, S., and Zhang, Y. (2015). Haploinsufficiency, but not defective paternal 5mC oxidation, accounts for the developmental defects of maternal Tet3 knockouts. *Cell Rep* 10, 463-470.
- Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabo, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A* 108, 3642-3647.
- Ishida, T., and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* 35, W460-464.
- Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., Sugawara, A., and Yokoyama, A. (2014). TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes Cells* 19, 52-65.
- Iurlaro, M., Ficz, G., Oxley, D., Raiber, E.A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S., and Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol* 14, R119.
- Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 8, 1698-1710.
- Janke, R., Dodson, A.E., and Rine, J. (2015). Metabolism and Epigenetics. *Annu Rev Cell Dev Biol* 10.1146/annurev-cellbio-100814-125544.
- Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science* 293, 1074-1080.
- Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429, 900-903.
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28, 27-30.
- Karve, T.M., and Cheema, A.K. (2011). Small changes huge impact: the role of protein posttranslational modifications in cellular homeostasis and disease. *J Amino Acids* 2011, 207691.
- Kim, D.I., Birendra, K.C., Zhu, W., Motamedchaboki, K., Doye, V., and Roux, K.J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc Natl Acad Sci U S A* 111, E2453-2461.

## References

- Ko, M., An, J., Bandukwala, H.S., Chavez, L., Aijo, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lahdesmaki, H., Hogan, P.G., Aravind, L., and Rao, A. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 497, 122-126.
- Konstandin, N., Bultmann, S., Szwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P.M., Schneider, S., Spiekermann, K., Leonhardt, H., and Bohlander, S.K. (2011). Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia* 25, 1649-1652.
- Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* 184, 868-871.
- Krokan, H.E., and Bjoras, M. (2013). Base excision repair. *Cold Spring Harb Perspect Biol* 5, a012583.
- Kustatscher, G., Hegarat, N., Wills, K.L., Furlan, C., Bukowski-Wills, J.C., Hochegger, H., and Rappsilber, J. (2014). Proteomics of a fuzzy organelle: interphase chromatin. *EMBO J* 33, 648-664.
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* 16, 2893-2905.
- Lazarus, M.B., Nam, Y., Jiang, J., Sliz, P., and Walker, S. (2011). Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* 469, 564-567.
- Lee, J.H., You, J., Dobrota, E., and Skalnik, D.G. (2010). Identification and characterization of a novel human PP1 phosphatase complex. *J Biol Chem* 285, 24466-24476.
- Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., Surani, M.A., and Hajkova, P. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat Struct Mol Biol* 20, 311-316.
- Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 71, 865-873.
- Liu, N., Wang, M., Deng, W., Schmidt, C.S., Qin, W., Leonhardt, H., and Spada, F. (2013). Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PLoS One* 8, e62755.
- Loenarz, C., and Schofield, C.J. (2009). Oxygenase catalyzed 5-methylcytosine hydroxylation. *Chem Biol* 16, 580-583.
- Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., and Downing, J.R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* 17, 637-641.
- Lothrop, A.P., Torres, M.P., and Fuchs, S.M. (2013). Deciphering post-translational modification codes. *FEBS Lett* 587, 1247-1257.
- Lu, C., and Thompson, C.B. (2012). Metabolic regulation of epigenetics. *Cell Metab* 16, 9-17.
- Lubas, W.A., Frank, D.W., Krause, M., and Hanover, J.A. (1997). O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *J Biol Chem* 272, 9316-9324.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251-260.
- Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286, 35334-35338.
- Mardis, E.R., Ding, L., Dooling, D.J., Larson, D.E., McLellan, M.D., Chen, K., Koboldt, D.C., Fulton, R.S., Delehaanty, K.D., McGrath, S.D., Fulton, L.A., Locke, D.P., Magrini, V.J., Abbott, R.M., Vickery, T.L., Reed, J.S., Robinson, J.S., Wylie, T., Smith, S.M., Carmichael, L., et al. (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 361, 1058-1066.
- Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J* 24, 800-812.
- McDonel, P., Costello, I., and Hendrich, B. (2009). Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. *Int J Biochem Cell Biol* 41, 108-116.

- Meissner, A. (2010). Epigenetic modifications in pluripotent and differentiated cells. *Nat Biotechnol* 28, 1079-1088.
- Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., Gnirke, A., Jaenisch, R., and Lander, E.S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766-770.
- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553-560.
- Minguez, P., Parca, L., Diella, F., Mende, D.R., Kumar, R., Helmer-Citterich, M., Gavin, A.C., van Noort, V., and Bork, P. (2012). Deciphering a global network of functionally associated post-translational modifications. *Mol Syst Biol* 8, 599.
- Morton, N.E. (1991). Parameters of the human genome. *Proc Natl Acad Sci U S A* 88, 7474-7476.
- Mulholland, C.B., Smets, M., Schmidtman, E., Leidescher, S., Markaki, Y., Hofweber, M., Qin, W., Manzo, M., Kremmer, E., Thanisch, K., Bauer, C., Rombaut, P., Herzog, F., Leonhardt, H., and Bultmann, S. (2015). A modular open platform for systematic functional studies under physiological conditions. *Nucleic Acids Res* 43, e112.
- Muller, I., Stuckl, C., Wakeley, J., Kertesz, M., and Uson, I. (2005). Succinate complex crystal structures of the alpha-ketoglutarate-dependent dioxygenase AtsK: steric aspects of enzyme self-hydroxylation. *J Biol Chem* 280, 5716-5723.
- Muller, U., Bauer, C., Siegl, M., Rottach, A., and Leonhardt, H. (2014). TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res* 42, 8592-8604.
- Muslin, A.J., Tanner, J.W., Allen, P.M., and Shaw, A.S. (1996). Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine. *Cell* 84, 889-897.
- Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol* 8, 751-758.
- Nakamura, T., Liu, Y.J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* 486, 415-419.
- Ngo, T.T., Yoo, J., Dai, Q., Zhang, Q., He, C., Aksimentiev, A., and Ha, T. (2016). Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat Commun* 7, 10813.
- Nichols, J., and Smith, A. (2012). Pluripotency in the Embryo and in Culture. *Cold Spring Harb Perspect Biol* 4, a008128.
- Nowak, S.J., and Corces, V.G. (2004). Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet* 20, 214-220.
- Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247-257.
- Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* 19, 219-220.
- Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). *Science* 183, 330-332.
- Olins, D.E., and Olins, A.L. (2003). Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* 4, 809-814.
- Palini, S., De Stefani, S., Scala, V., Dusi, L., and Bulletti, C. (2011). Epigenetic regulatory mechanisms during preimplantation embryo development. *Ann N Y Acad Sci* 1221, 54-60.
- Parsons, D.W., Jones, S., Zhang, X., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.M., Gallia, G.L., Olivi, A., McLendon, R., Rasheed, B.A., Keir, S., Nikolskaya, T., Nikolsky, Y., Busam, D.A., Tekleab, H., Diaz,



## References

- L.A., Jr., Hartigan, J., *et al.* (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science* 321, 1807-1812.
- Patel, J., Pathak, R.R., and Mujtaba, S. (2011). The biology of lysine acetylation integrates transcriptional programming and metabolism. *Nutr Metab (Lond)* 8, 12.
- Perteau, M., and Salzberg, S.L. (2010). Between a chicken and a grape: estimating the number of human genes. *Genome Biol* 11, 206.
- Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S.K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., Schuermann, D., Michalakis, S., Kosmatchev, O., Schiesser, S., Steigenberger, B., Raddaoui, N., Kashiwazaki, G., Muller, U., Spruijt, C.G., Vermeulen, M., *et al.* (2014). Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol* 10, 574-581.
- Qin, W., Leonhardt, H., and Spada, F. (2011). Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J Cell Biochem* 112, 439-444.
- Qin, W., Wolf, P., Liu, N., Link, S., Smets, M., Mastra, F.L., Forne, I., Pichler, G., Horl, D., Fellingner, K., Spada, F., Bonapace, I.M., Imhof, A., Harz, H., and Leonhardt, H. (2015). DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res* 25, 911-929.
- Rayasam, G.V., Wendling, O., Angrand, P.O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G.L., Chambon, P., and Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J* 22, 3153-3163.
- Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. *Science* 293, 1089-1093.
- Reither, S., Li, F., Gowher, H., and Jeltsch, A. (2003). Catalytic Mechanism of DNA-(cytosine-C5)-methyltransferases Revisited: Covalent Intermediate Formation is not Essential for Methyl Group Transfer by the Murine Dnmt3a Enzyme. *J Mol Biol* 329, 675-684.
- Richards, E.J., and Elgin, S.C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* 108, 489-500.
- Rottach, A., Frauer, C., Pichler, G., Bonapace, I.M., Spada, F., and Leonhardt, H. (2010). The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res* 38, 1796-1804.
- Rottach, A., Leonhardt, H., and Spada, F. (2009). DNA methylation-mediated epigenetic control. *J Cell Biochem* 108, 43-51.
- Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* 196, 801-810.
- Santi, D.V., Garrett, C.E., and Barr, P.J. (1983). On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell* 33, 9-10.
- Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C., and Leonhardt, H. (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res* 35, 4301-4312.
- Schomacher, L., Han, D., Musheev, M.U., Arab, K., Kienhofer, S., von Seggern, A., and Niehrs, C. (2016). Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nat Struct Mol Biol* 23, 116-124.
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell* 128, 735-745.
- Schwartz, M.W., Baskin, D.G., Bukowski, T.R., Kuijper, J.L., Foster, D., Lasser, G., Prunkard, D.E., Porte, D., Jr., Woods, S.C., Seeley, R.J., and Weigle, D.S. (1996). Specificity of leptin action on elevated blood glucose levels and hypothalamic neuropeptide Y gene expression in ob/ob mice. *Diabetes* 45, 531-535.
- Seet, B.T., Dikic, I., Zhou, M.M., and Pawson, T. (2006). Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* 7, 473-483.
- Seo, J., and Lee, K.J. (2004). Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *J Biochem Mol Biol* 37, 35-44.

- Shafi, R., Iyer, S.P., Ellies, L.G., O'Donnell, N., Marek, K.W., Chui, D., Hart, G.W., and Marth, J.D. (2000). The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc Natl Acad Sci U S A* 97, 5735-5739.
- Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem* 76, 75-100.
- Simmons, D.G., and Cross, J.C. (2005). Determinants of trophoblast lineage and cell subtype specification in the mouse placenta. *Dev Biol* 284, 12-24.
- Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* 336, 688-690.
- Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* 484, 339-344.
- Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat Rev Genet* 14, 204-220.
- Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* 331, 1036-1040.
- Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F., Eberl, H.C., Mensinga, A., Brinkman, A.B., Lephikova, K., Muller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., Carell, T., *et al.* (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146-1159.
- Stec, I., Nagl, S.B., van Ommen, G.J., and den Dunnen, J.T. (2000). The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? *FEBS Lett* 473, 1-5.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930-935.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663-676.
- Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* 14, 1025-1040.
- Ubersax, J.A., and Ferrell, J.E., Jr. (2007). Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* 8, 530-541.
- Uhart, M., and Bustos, D.M. (2014). Protein intrinsic disorder and network connectivity. The case of 14-3-3 proteins. *Frontiers in Genetics* 5, 1-7.
- UniProt, C. (2015). UniProt: a hub for protein information. *Nucleic Acids Res* 43, D204-212.
- Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., and Pasini, D. (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol Cell* 49, 645-656.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., *et al.* (2001). The sequence of the human genome. *Science* 291, 1304-1351.
- Voncken, J.W., Roelen, B.A., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J., and van Lohuizen, M. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc Natl Acad Sci U S A* 100, 2468-2473.
- Waddington, C.H. (1939). Development as an Epigenetic Process. In: *Introduction to Modern Genetics*, New York: The Macmillan Company, pp. 154-156.
- Wang, H., and Dey, S.K. (2006). Roadmap to embryo implantation: clues from mouse models. *Nat Rev Genet* 7, 185-199.

## References

- Wang, Z., Gucek, M., and Hart, G.W. (2008). Cross-talk between GlcNAcylation and phosphorylation: site-specific phosphorylation dynamics in response to globally elevated O-GlcNAc. *Proc Natl Acad Sci U S A* 105, 13793-13798.
- Wang, Z., Pandey, A., and Hart, G.W. (2007). Dynamic interplay between O-linked N-acetylglucosaminylation and glycogen synthase kinase-3-dependent phosphorylation. *Mol Cell Proteomics* 6, 1365-1379.
- Weber, A.R., Krawczyk, C., Robertson, A.B., Kusnierczyk, A., Vagbo, C.B., Schuermann, D., Klungland, A., and Schar, P. (2016). Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat Commun* 7, 10806.
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343-348.
- Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* 336, 684-687.
- Wolffe, A.P., and Guschin, D. (2000). Review: chromatin structural features and targets that regulate transcription. *J Struct Biol* 129, 102-122.
- Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* 2, 241.
- Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473, 389-393.
- Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Wang, P., Xiao, M.T., Liu, L.X., Jiang, W.Q., Liu, J., Zhang, J.Y., Wang, B., Frye, S., Zhang, Y., Xu, Y.H., Lei, Q.Y., Guan, K.L., *et al.* (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* 19, 17-30.
- Yaffe, M.B., Rittinger, K., Volinia, S., Caron, P.R., Aitken, A., Leffers, H., Gamblin, S.J., Smerdon, S.J., and Cantley, L.C. (1997). The structural basis for 14-3-3:phosphopeptide binding specificity. *Cell* 91, 961-971.
- Yan, H., Parsons, D.W., Jin, G., McLendon, R., Rasheed, B.A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G.J., Friedman, H., Friedman, A., Reardon, D., Herndon, J., Kinzler, K.W., Velculescu, V.E., Vogelstein, B., and Bigner, D.D. (2009). IDH1 and IDH2 mutations in gliomas. *N Engl J Med* 360, 765-773.
- Yin, R., Mao, S.Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., Jiao, Y., Li, C., Liu, S., Wu, D., Gu, W., Yang, Y.G., Xu, G.L., and Wang, H. (2013). Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J Am Chem Soc* 135, 10396-10403.
- Ying, Q.L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519-523.
- Zhang, Q., Liu, X., Gao, W., Li, P., Hou, J., Li, J., and Wong, J. (2014). Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked beta-N-acetylglucosamine transferase (OGT). *J Biol Chem* 289, 5986-5996.
- Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev* 15, 2343-2360.

## Annex

### Abbreviations

2-HG	2-hydroxyglutarate
2i	2 inhibitors (PD0325901 and CHIR99021)
2-OG	2-oxoglutarate
ADD	ATRX/DNMT3/DNMT3L
AML	acute myeloid leukemia
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BAH	bromo-adjacent homology
BER	base excision repair
bp	base pairs
caC	carboxylcytosine
CC	coiled-coil
CGIs	CpG islands
ChEP	chromatin enrichment for proteomics
co-IP	co-immunoprecipitation
CTD	C-terminal domain
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
DPPA3	Developmental pluripotency-associated protein 3
DSB	double strand break
EGF	epidermal growth factor
EMSA	electrophoretic mobility shift assay
EpiLC	epiblast-like cell
ESC	embryonic stem cell
EZH2	enhancer of zeste homolog 2
F3H	fluorescent-3-hybrid
fC	formylcytosine
FDR	false-discovery-rate
FGF2	fibroblast growth factor 2
GFP	green fluorescent protein
GO	gene ontology
GSK3	Glycogen synthase kinase-3
HAT	histone acetyltransferase
HCFC1	host cell factor 1
HDAC	histone deacetylase
HeLa	Henrietta Lacks (human epitheloid cervix carcinoma cell line)
hmC	hydroxymethylcytosine
hmU	hydroxymethyluracil
IAP	intracisternal-A particles
ICM	inner cell mass
IDH	isocitrate dehydrogenase
IP	Immunoprecipitation
KD	knockdown
KO	knockout
LC-MS/MS	liquid chromatography tandem mass spectrometry
LFQ	label-free quantification
LIF	Leukemia inhibitory factor
LINE	long interspersed nuclear elements
LTR	long terminal repeat
MAPK	mitogen-activated protein kinase
MBD	methyl CpG binding domain
mC	methylcytosine

## Annex

MEF	mouse embryonic fibroblast
MLL	Mixed lineage leukemia
MPG	methylpurine DNA glycosylase
NA	numeric aperture
NAD(P)+	nicotinamide adenine dinucleotide (phosphate)
NAD(P)H	reduced NAD(P)+
NEIL	Nei-like protein (or Endonuclease 8-like protein)
NGF	nerve growth factor
OGA	O-GlcNAcase
O-GlcNAc	O-linked N-Acetylglucosamine
OGT	O-GlcNAc transferase
p.f.	post fertilization
PCA	principal component analysis
PCNA	proliferating cell nuclear antigen
PGC	primordial germ cell
PHD	plant homeodomain
PRC	polycomb repressive complex
PTM	post-translational modification
PWWP	proline-tryptophan-tryptophan-proline
RNA	ribonucleic acid
RT	room temperature
SAH	S-Adenosyl-homocysteine
SAM	S-Adenosylmethionine
SET	Su(var)3-9/Enhancer-of-zeste/Trithorax
SILAC	stable isotope labeling by amino acids in cell culture
SRA	SET and RING associated
SUZ12	suppressor of zeste 12 protein homolog
TCA	tricarboxylic acid
TDG	Thymine-DNA glycosylase
TE	trophoblast
TET	ten-eleven-translocation
TKO	triple knockout
TNF- $\alpha$	tumor necrosis factor $\alpha$
TPR	tetratricopeptide repeat
TRD	transcriptional repression domain
TSS	transcription start sites
TTD	tandem tudor domain
UDP-GlcNAc	uridine diphosphate N-Acetylglucosamine
UHRF	Ubiquitin-like PHD and RING finger domain-containing protein
UIM	ubiquitin interaction motif
wt	wild type

## List of publications

Mulholland, C.B., Smets, M., Schmidtman, E., Leidescher, S., Markaki, Y., Hofweber, M., Qin, W., Manzo, M., Kremmer, E., Thanisch, K., **Bauer, C.**, Rombaut, P., Herzog, F., Leonhardt, H., and Bultmann, S. (2015). A modular open platform for systematic functional studies under physiological conditions. *Nucleic acids research* 43, e112.

**Bauer, C.**, Gobel, K., Nagaraj, N., Colantuoni, C., Wang, M., Muller, U., Kremmer, E., Rottach, A., and Leonhardt, H. (2015). Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *The Journal of biological chemistry* 290, 4801-4812.

Muller, U., **Bauer, C.**, Siegl, M., Rottach, A., and Leonhardt, H. (2014). TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic acids research* 42, 8592-8604.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., **Bauer, C.**, Munzel, M., Wagner, M., Muller, M., Khan, F., Eberl, H.C., Mensinga, A., Brinkman, A.B., Lephikov, K., Muller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., Carell, T., and Vermeulen M. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146-1159.

## **Statutory declaration and statement**

### **Eidesstattliche Erklärung**

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt wurde.

München, den 12.04.2016

Christina Bauer

### **Erklärung**

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig ohne Erfolg einer Doktorprüfung unterzogen habe.

München, den 12.04.2016

Christina Bauer



## Acknowledgements

First, I want to thank my Doktorvater Heinrich Leonhardt for his constant support, not only during my time as a doctoral student, but also before during my Master's program. I especially want to thank him for encouraging the attendance of conferences and courses, thereby broadening my personal and professional horizon.

I furthermore want to thank my two collaborators Michiel Vermeulen and Nagarjuna Nagaraj, from whom I learned a lot about mass spectrometry and proteomics and who always offered advice when needed. I also want to thank all of my colleagues at the Leonhardt Lab, both past and present, for providing a great working atmosphere. Special thanks go to some colleagues in particular:

To Udo Müller, for first being a great Master's thesis supervisor, and later a fantastic colleague and office mate. We always very fruitfully cooperated on our common projects, nicely complementing each other's strengths and weaknesses. We also had matching ideas about how to spend a free evening, be it at the Paulaner Biergarten or the Friedrichstadtpalast in Berlin. Thank you for everything and I wish you all the best for the future!

To Chris Mulholland and Yolanda Markaki, for a million inspiring scientific discussions, both on and off duty. Additional thanks go to Chris for a lot of proof-reading and suggestions on manuscripts. On a more personal note, I will miss both of you a lot, above all the going out for dinner and dancing parts, but also the debates about science, career, politics, family, and everything else. I sincerely hope we will stay in touch.

To Elisabeth Schmidtman, for constant discussions about proteomics and data analysis, and for a fantastic time in Washington DC. It was great having you as an office neighbor.

To everybody else who provided ideas, insights, and technical advice on various projects and experiments: Martha Smets, Joel Ryan, Katharina Thanisch, Patricia Wolf, Andrea Rottach, Sebastian Bultmann, Carina Frauer, Garwin Pichler, Jonas Helma, and Irina Solovei.

To Danny Meilinger for shared teaching Ups and Downs – “Think first” was a definite Up! Additional thanks for lunch conversations, party hosting, TV sharing, and sleep-over opportunities; we had a great time together.

Lastly, I want to express my deepest thanks to my family. To my grandparents, aunts, uncles, and cousins for just being who they are and for providing a completely different down-to-earth perspective on so many aspects of life and work. To my parents Susanne and Johannes who have always supported me to find my own way and make my own decisions. Thank you for always helping me with practical and emotional issues, and for everything else I cannot put into words here. To my sister Gabriela who is the best and most reliable friend I have. And to my husband Robert, who always believes in me.